

平成 25 年度修士論文

視線ジェスチャによる  
携帯デバイスのページ操作支援

学籍番号 1230052

氏名 竹内 尚也

総合情報学専攻 メディア情報学コース

指導教員 高橋 裕樹 准教授 印

副指導教員 坂本 真樹 准教授 印

提出日 平成 26 年 1 月 30 日 (木)



# 概要

タブレット PC を使用するときの問題として、両手が塞がった場合、機器の操作が困難になる。この問題を解決するために、手を用いない代替操作手法として視線操作が挙げられる。しかし、タブレット PC での視線操作には次の 3 つの問題点がある。(1) タブレット PC の画面が小さいため、高い視線計測が必要。(2) ユーザとタブレット PC の位置関係が変化しやすく、注視点がずれやすい。(3) 正確な視線検出を行う特別な装置を付けるとタブレット PC の携帯性が失われてしまう。作業しながら本などを閲覧するときは、ページをめくるコマンド操作が重要となる。視線ジェスチャは限られた時間内に行われる様々な眼球運動のパターンであり、コマンド操作を視線ジェスチャで代替することができる。本論文では、視線ジェスチャを用いた携帯デバイスのページ操作支援を提案した。

まず、タブレット PC に搭載されている単眼の Web カメラを用いて画像処理による視線検出を行った。視線は眼球中心と虹彩中心を結んだベクトルとする。顔に対して AAM(Active Appearance Model) を用いて特徴点追跡を行う。眼球中心は眼球付近の特徴点を用いて求められる。虹彩中心は眼球領域を 2 値化して得られる黒色領域に楕円を近似して求める。画像からは虹彩の偏りしか求められないため目の幅を用いて眼球半径を推定し、視線ベクトルを求めた。その後、キャリブレーションを行うことで注視点を推定した。次に提案システムで注視点の計測実験を行った。鉛直方向は水平方向と比べて平均誤差と標準偏差が大きく、正確な注視点が安定して得られていないが、視線ジェスチャの識別が可能な計測制度を得られた。

注視点の平均誤差から、注視点の軌跡による視線ジェスチャを定義した。識別精度が低く、未識別も高かったため、領域移動の遷移を用いた視線ジェスチャ識別手法を提案した。これにより精度は向上し、未識別が少なくなった。注視点を用いた視線ジェスチャは頭部が移動すると正確な注視点が得られず、視線ジェスチャが識別できない問題がある。そのため、視線ベクトルによる視線ジェスチャを定義した。視線ベクトルの角度変化はタブレット PC とユーザの距離によって異なるため、タブレット PC からユーザ頭部までの距離を測定した。その結果からサッカードとなる眼球運動を定義し、水平方向と鉛直方向のサッカードの組み合わせから視線ジェスチャを識別した。サッカードの検出はフレーム間差分による手法と最小 2 乗法による手法を用いた。フレーム間差分による手法は誤差の影響が大きいため、視線ジェスチャの識別精度が低く、未識別も高くなった。最小 2 乗法による手法による手法は識別精度が高く、誤識別と未識別は低くなり、注視点の軌跡による手法よりも高精度な視線ジェスチャ識別が可能となった。しかし、読書時の意図しない視線ジェスチャの回数は、他の手法と比べて多かった。今後、頭部を固定せずに視線ジェスチャ識別の精度評価の検討を行う必要がある。また、読書時に意図しない視線ジェスチャが識別されないように、より適切な視線ジェスチャの動きや時間の制約を検討する。



# 目 次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	背景と目的 . . . . .	1
1.2	本論文の構成 . . . . .	2
<b>第 2 章</b>	<b>従来研究</b>	<b>3</b>
2.1	視線検出手法 . . . . .	3
2.1.1	特別な装置を用いた手法 . . . . .	4
2.1.2	画像処理による手法 . . . . .	6
2.2	視線を用いたインタフェース . . . . .	10
2.2.1	注視点の停留時間の利用 . . . . .	10
2.2.2	注視点の軌跡の利用 . . . . .	12
<b>第 3 章</b>	<b>視軸推定</b>	<b>15</b>
3.1	特徴点追跡 . . . . .	15
3.2	虹彩中心の検出 . . . . .	17
3.3	視線ベクトル取得 . . . . .	18
3.4	注視点推定 . . . . .	21
3.5	評価実験 . . . . .	21
3.5.1	被験者 . . . . .	21
3.5.2	実験手順 . . . . .	21
3.5.3	実験結果 . . . . .	24
<b>第 4 章</b>	<b>注視点による視線ジェスチャ</b>	<b>27</b>
4.1	ジェスチャの定義 . . . . .	27
4.2	視線ジェスチャによる書籍の操作 . . . . .	27
4.3	領域移動の遷移による視線ジェスチャの識別 . . . . .	28
4.4	評価実験 . . . . .	30
<b>第 5 章</b>	<b>視線ベクトルによる視線ジェスチャ</b>	<b>31</b>
5.1	タブレット PC とユーザ頭部の距離測定 . . . . .	31
5.2	視線ジェスチャの定義 . . . . .	33
5.2.1	フレーム間差分によるサッカード検出 . . . . .	34
5.2.2	最小 2 乗法によるサッカード検出 . . . . .	36

5.3	視線ジェスチャの識別 . . . . .	36
5.4	評価実験 . . . . .	39
第 6 章	結論	41
6.1	本論文の結論 . . . . .	41
6.2	今後の課題 . . . . .	41
謝辞		43

# 目 次

2.1	眼球の構造	3
2.2	眼球画像の比較 [1]	4
2.3	眼球周辺の電位差	4
2.4	サーチコイル法 [9]	5
2.5	強膜反射法	5
2.6	プルキニエ像 [1]	6
2.7	画像処理による視線検出手法	6
2.8	情報提示システムの外観 [12]	7
2.9	虹彩モデル [12]	7
2.10	Lucas-Kanade の特徴点追跡 [14]	8
2.11	眼球モデルと観測画像の誤差 [14]	8
2.12	AAM の特徴点追跡 [16]	9
2.13	コンピュータ操作支援システムの指標配置 [23]	11
2.14	注視状態判定による入力システム	11
2.15	実験の作業風景 [24]	11
2.16	注視領域と視線ジェスチャ [25]	12
2.17	視線ジェスチャによるキャラクターの操作 [25]	13
2.18	視線ジェスチャの方向 [27]	13
2.19	視線ジェスチャによるブラウザ操作コマンド [27]	14
2.20	実験環境 [27]	14
3.1	提案手法の処理概要	16
3.2	顔の特徴点追跡	17
3.3	眼球の特徴点	17
3.4	虹彩抽出	18
3.5	眼球モデル	20
3.6	目幅と眼球半径	20
3.7	注視点推定	22
3.8	キャリブレーション	22
3.9	実験の様子	23
3.10	注視ターゲット	23
3.11	注視点の平均誤差	25

---

4.1	注視点の軌跡を用いた視線ジェスチャによるページ操作コマンド . . . . .	27
4.2	高速な視線ジェスチャ . . . . .	28
4.3	領域遷移による視線ジェスチャ識別 . . . . .	29
4.4	視線ジェスチャ識別の失敗 . . . . .	29
5.1	Kinect®による測定の様子 . . . . .	32
5.2	骨格追跡 . . . . .	32
5.3	視線角度変化を用いた視線ジェスチャによるページ操作コマンド . . . . .	33
5.4	水平方向のサッカード . . . . .	34
5.5	フレーム間差分によるサッカード検出 . . . . .	35
5.6	優先されるサッカード . . . . .	35
5.7	最小2乗法によるサッカード検出 . . . . .	36
5.8	視線ジェスチャの遷移 . . . . .	37
5.9	サッカードの遷移による視線ジェスチャの識別 . . . . .	38
5.10	ジェスチャ識別後 . . . . .	38



# 表 目 次

2.1	視線検出手法の比較 . . . . .	9
3.1	注視点の平均誤差と標準偏差 ( $mm$ ) . . . . .	24
4.1	注視点の軌跡による視線ジェスチャの識別精度 (%) . . . . .	28
4.2	領域移動の遷移による視線ジェスチャの識別精度 (%) . . . . .	30
4.3	意図しない注視点の軌跡による視線ジェスチャ(回) . . . . .	30
5.1	頭部への距離の平均誤差と標準偏差 ( $mm$ ) . . . . .	31
5.2	視線角度変化による視線ジェスチャの識別精度 (%) . . . . .	39
5.3	意図しない視線角度変化による視線ジェスチャ(回) . . . . .	39

# 第1章 序論

## 1.1 背景と目的

現在、様々な携帯デバイスが開発されている。電話、インターネット、電子書籍など、異なるデバイスが扱っていた情報が扱える新しいデバイスとして注目されている。特に iPad® などに代表されるタブレット PC(Personal Computer) は、高性能と携帯性を兼ね備えた端末である。大きな表示画面を持ち、見る角度に依存しない、紙媒体に近いデバイスである。雑誌や新聞、書籍、楽譜、台本など、紙媒体で利用されていたものが、タブレット PC で利用されるようになってきている。タッチスクリーンに対する直感的な操作で、誰もが容易に操作できる。しかし、タブレット PC を使用するときの問題として、両手が塞がり操作が困難になる場合がある。例えば、電車で立っている状態では、片方の手でつり革を掴み、もう片方の手でタブレット PC を支える必要がある。また、タブレット PC でレシピを見ながら料理を作る際は、両手を調理に使うためタブレット PC の操作が行えない。この問題を解決するために、両手に依存しない代替操作手法として視線操作 [1] が挙げられる。

視線操作には重要な問題として “Midas Touch Problem” [2] がある。これはユーザが視線操作のために意図的に見ているのか、意図せずに見ているかが区別できないことである。例えば、PC においてアイコンを選択する場合、まず対象のアイコンを見つめ、マウス操作などでポインタをアイコンまで移動する必要がある。もし、アイコンを見つめるだけで選択できるのなら、従来のインタフェースより高速で直感的にアイコン選択が行える。しかし、アイコンを見るだけで選択されてしまうため、多くのアイコンの中から目的のアイコンを探す場合、見たアイコンが全て選択されてしまい不便である。

また、タブレット PC など携帯デバイスで視線操作を行うには 3 つの問題点がある。1 つ目は、タブレット PC の画面とアイコンが小さいことである。視線操作ではユーザの正確な注視点が要求される。注視点の誤差がアイコンのサイズより大きいと、あるアイコンを見つめても、隣のアイコンを見つめていると誤検出されることがある。2 つ目は、ユーザの顔を固定できないことである。タブレット PC を利用する際は手で支えることがあり、ユーザとタブレット PC の位置関係が変化しやすい。そのため、ユーザのタブレット PC への注視点がずれてしまいキャリブレーションなどをやり直す必要がある。3 つ目は、正確な視線検出には特別な装置が必要なことである。例えば一般的な視線検出に用いられる角膜反射法 [3] は、赤外線光と赤外線カメラが必要となる。このような装置を付けるとタブレット PC の携帯性が失われてしまう。

作業しながら本、レシピ、楽譜を閲覧するときは、マウスやキーボードによる複雑な操作よりも、ページをめくるなどのコマンド操作が重要となる。視線ジェスチャ [4] は限られた時

間内で行われる様々な眼球運動パターンであり，コマンド操作を視線ジェスチャで意図することができる．視線ジェスチャは“どこを見ているか”ではなく“どのように見ているか”が重要となるため正確な注視点は必要としない．そのため本論文では，視線ジェスチャを用いた携帯デバイスのページ操作支援を検討する．

## 1.2 本論文の構成

本論文は序論にて，背景と目的を述べた．第2章では，まず様々な視線検出手法について述べる．特に可視光画像を用いた視線検出手法について詳しく述べる．その後，視線を用いたインタフェースについて，注視点の停留時間を用いた手法と注視点の軌跡を用いた手法を述べる．第3章では，眼球のモデルについて説明し，画像から人物の視線ベクトルを取得し，キャリブレーションによって注視点を推定する方法を示し，評価実験を行う．第4章では，注視点を用いた視線ジェスチャ操作の検討と評価実験を行う．第5章では，視線ベクトルを用いた視線ジェスチャ操作の検討と評価実験を行う．第6章では，本論文の結論と今後の課題を述べる．

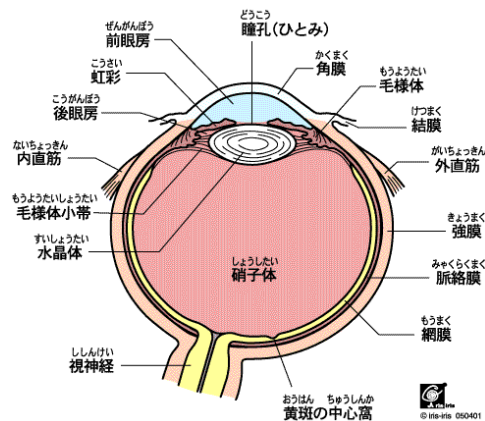
## 第2章 従来研究

視線検出手法は特別な装置を用いた手法と画像処理による手法に大別できる．画像処理による手法はカメラのみで視線検出が行える．本論文ではタブレット PC のみで視線操作を行うことを目的としてゐるため，画像処理による手法で視線検出を行う．

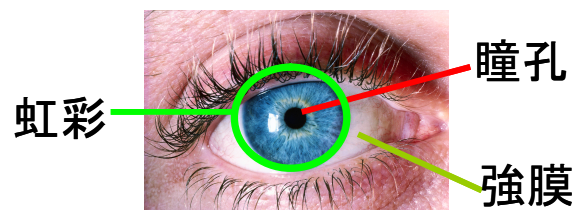
### 2.1 視線検出手法

眼球の構造を図 2.1 に示す．図 2.1(a) は眼球の断面図であり，図 2.1(b) は正面から見た眼球である．眼球の動きや瞳孔の向きを調べることで視線検出が行える．図 2.2 は眼球の可視光と赤外線画像の比較である．図 2.2(a) に示すように個人によって可視光画像における瞳孔は虹彩との区別が難しい．そのため瞳孔の向きから視線検出を行う手法では，図 2.2(b) に示す赤外線画像を用いて瞳孔を検出することが一般的である．視線検出手法には，EOG(Electro-oculogram) 法，サーチコイル法，強膜反射法/強膜トラッカー法，角膜反射法，画像処理を用いた手法などがある．

め こうぞう  
■目の構造（右眼球の水平断）



(a) 眼球の解剖図 [5]



(b) 正面から見た眼球

図 2.1: 眼球の構造

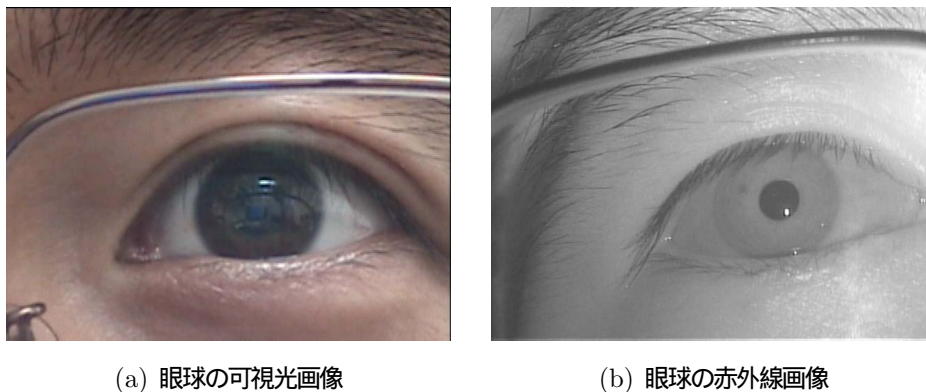


図 2.2: 眼球画像の比較 [1]

### 2.1.1 特別な装置を用いた手法

EOG 法 [6] は眼球を動かしたときに生じる目周辺の皮膚の電位差を利用することで眼球の運動を測定し、視線を推定する手法である。眼球は図 2.1(a) に示す構造をしており、角膜は眼球後部の網膜に対して、正に帯電している。そのため図 2.3 で示すように眼球の動きによって目周辺で電位が変化する。角膜と網膜の電位差は筋肉や眼球への光の強さの影響を受けるため、視線検出の精度は高くない [7]。しかし簡易な装置であり、皮膚電極を貼り付けるだけであり利用者の負担には殆どならない。また、顔を固定せずに視線検出が可能である。眼球を閉じた状態でも使用できるため、睡眠時における眼球運動の調査 [8] に使用されている。

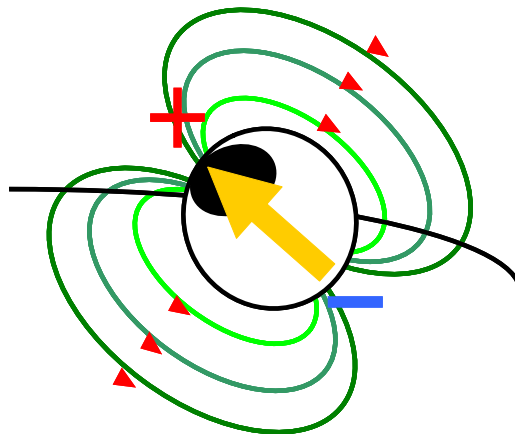


図 2.3: 眼球周辺の電位差

サーチコイル法 [9] はコンタクトレンズの外側にコイルを組み込み、コイルの位置を測定することで、瞳の動きを検出し視線を求める手法である。ドーナツ型のコンタクトレンズ上に円形にコイルを配置し、中央は瞳孔が隠れないように図 2.4 に示すように穴が開いている。磁場発生装置の中で頭部を固定する必要があるが、眼球の動きが直接分かり眼球を閉じた状態でも測定が行える高精度な視線検出となっている。正確な検出のためサーチコイル法

では，コンタクトレンズと眼球が密着している必要がある．コイルを密着させるのは痛みが生じ，その痛みを軽減するために点眼用の麻酔を使用する．そのため不快感があり利用者への負担は非常に大きく，長時間使用することは困難である．人への使用は殆ど行われず，動物の視線検出 [10] に多く用いられている．

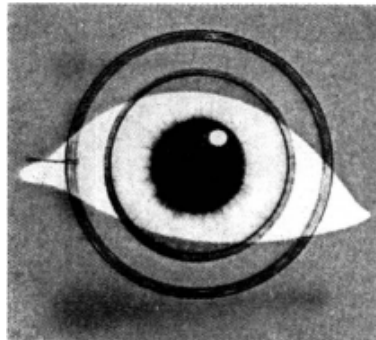


図 2.4: サーチコイル法 [9]

強膜反射法 [11] は眼球の黒目部分の角膜と白目部分の強膜の光の反射強度の違いを利用することで，視線検出を行う手法である．反射強度の違いを分かりやすくするために，反射光は眼球に近赤外線を照射して生じさせる．フォトダイオードなどの受光素子を用いて生じた反射光を検出する．角膜と強膜では，角膜の反射光のほうが強く，角膜は眼球の運動に伴い変化するため，図 2.5 に示すようにその差異を用いて眼球運動を検出する．水平方向の視線検出率は高いが，垂直方向ではあまり精度が期待できない．これは顔における眼球の形状が水平方向に長く，垂直方向は短いためである．また，受光素子の位置は固定されており，眼球から離れると測定が困難となるため，頭部に装置を装着する工夫や，頭部と装置を固定する工夫が必要になる．しかし，必要な装置は小型で軽量という利点がある．

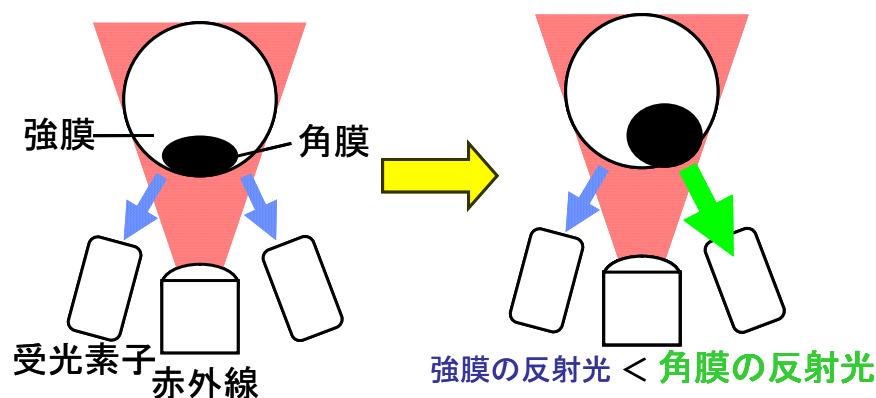


図 2.5: 強膜反射法

角膜反射法 [3] は，図 2.6 に示すように近赤外線を眼球に照射し角膜表面で反射するプルキニエ像をカメラで撮影することで，視線を検出する手法である．角膜の曲率中心と瞳孔中心を結んだ線が視線となる．角膜の曲率中心は，プルキニエ像を用いて算出する．瞳孔中心は

近赤外線画像から，楕円中心や二値化を用いて取得する．そのため視線を得るために頭部に装置を付ける場合もある．一方で視線検出する対象がパソコンのディスプレイ画面のように位置が固定されている場合，カメラを利用者の眼球が撮影可能な位置に固定することで，装置の装着をせずに視線検出が可能である．利用者の負担が少なく，精度の高い視線検出が可能のため，市販されている視線検出機器の多くは角膜反射法を用いている．

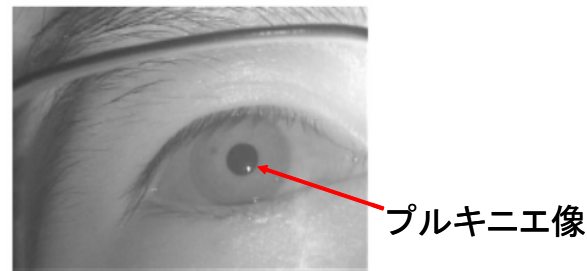


図 2.6: プルキニエ像 [1]

### 2.1.2 画像処理による手法

画像処理による手法は眼球をカメラで撮影し，虹彩の偏りから視線を検出する手法である．処理概要を図 2.7 に示す．顔全体を撮影し，その画像から目領域を求める．そのために特徴点追跡によって事前に目や鼻など顔のパーツの位置を取得する．目領域に 2 値化やハフ変換を用いることで虹彩の抽出を行う．その後，眼球特徴点と抽出された虹彩を眼球モデルの眼球領域中心と虹彩中心へ対応させ，視線を求める．カメラだけで視線検出可能だが精度は高くない．また可視光画像を用いるため，照明条件や肌，虹彩の色などの影響を受ける．しかし角膜反射法よりも制約が少なく，より広い範囲で簡易な視線検出が可能である．また視線検出の過程で顔全体を撮るため，顔の動きや表情も取得することができる．

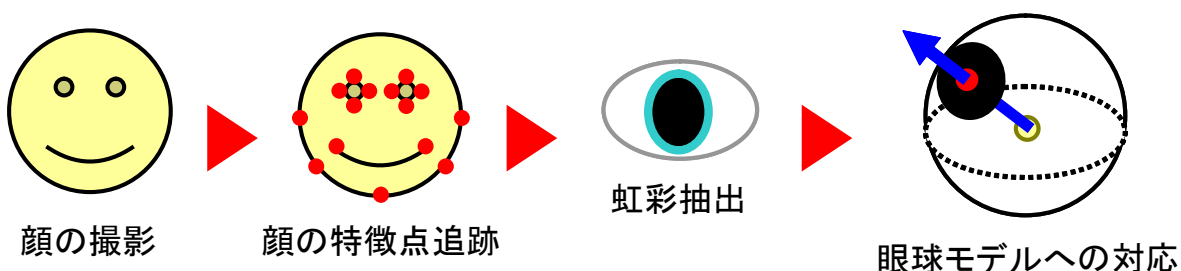


図 2.7: 画像処理による視線検出手法

佐竹ら [12] は複数の高解像度カメラで人物の視線を検出し，視線から人物の興味や反応を認識して，図 2.8 に示すインタラクティブな情報提示を行うシステムを提案した．装置を装着する必要がなく，ディスプレイに近づくだけで視線が検出できる．高解像度カメラで人物



の顔を撮影し，顔の特徴点を AAM(Active Appearance Model)[13] を用いて追跡することで，眼球の位置や顔の向きを得る．ステレオカメラで事前に学習することで，特徴点の 3 次元座標を求めることができる．そのため顔の向き変化を許容した視線検出が可能となっている．虹彩は図 2.9 に示すように球面形状を持たせてモデル化することで，実画像における見え方から視線推定に必要な虹彩モデルのパラメータを求めている．眼球領域を白色，黒色，眼球外領域に分け虹彩モデルを用いることで，虹彩中心を算出する．視線を眼球の旋回中心と虹彩中心を結ぶ線と定義し，各カメラで独立した視線が得られる．そこから顔特徴点や虹彩追跡の評価が低いものを除き，残った視線情報の重み付き平均によって正しい視線を推定する．視線の平均誤差は約 3 度であり，精度は十分に高いとは言えないが顔の特徴点を追跡しているため発展として表情の利用が行える．意図的な操作は行えないが，ディスプレイに対して一定の範囲で顔の向きに関係なく視線検出が可能である．

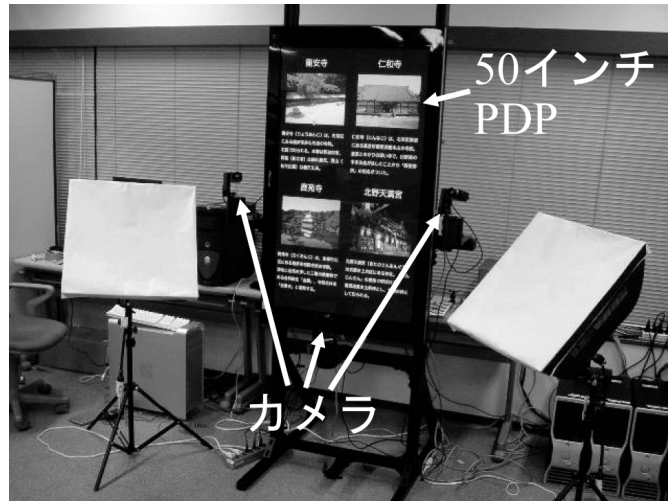


図 2.8: 情報提示システムの外観 [12]

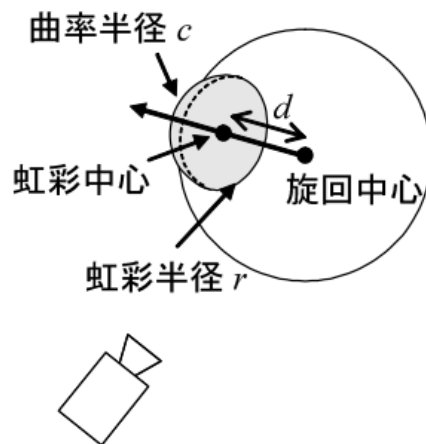


図 2.9: 虹彩モデル [12]



山添ら [14] は単眼カメラのみを用いた視線推定システムを提案している．従来の視線推定手法のもつ制約が緩和しており，自動のキャリブレーションで視線検出が行える．図 2.10 に示すように Lucas-Kanade の特徴点追跡 [15] によって得られた点群から，因子分解法を用いることで顔特徴点の 3 次元モデルを推定する．視線を眼球中心と虹彩中心を結ぶ 3 次元ベクトルと仮定することで，虹彩と白目の眼球モデルから視線推定を行っている．虹彩中心は画像列から求められるが，眼球中心と眼球半径は画像からは分からない．図 2.11 に示すように複数フレームにおける観測画像と，虹彩と白目のアピランスをもった 3 次元眼球モデルの誤差を最小にすることで眼球のパラメータを求めている．向きの異なる複数の顔画像があれば視線推定が行えるため，特別なキャリブレーションを行わず，最初に顔を自由に動かすだけでよい．十分に高解像度の顔画像が得られれば，距離に関係なくユーザの視線方向が検出できる．視線の平均誤差は約 5 度であり，特に虹彩が下部や端などに大きく偏っているときに視線精度が低くなっている．これは虹彩の偏りが大きいと瞼などによる隠蔽される虹彩領域が大きくなるためである．

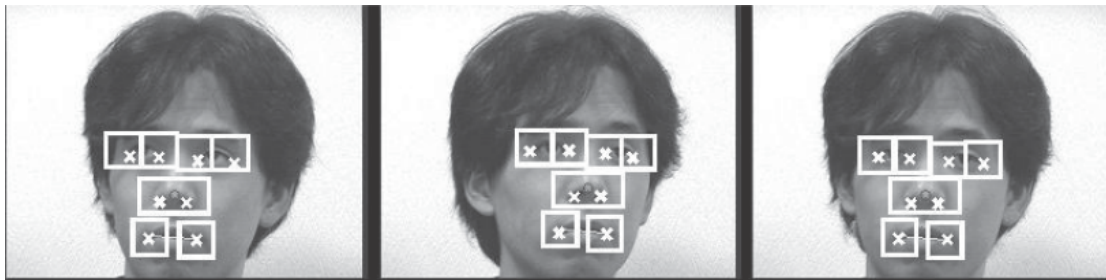


図 2.10: Lucas-Kanade の特徴点追跡 [14]

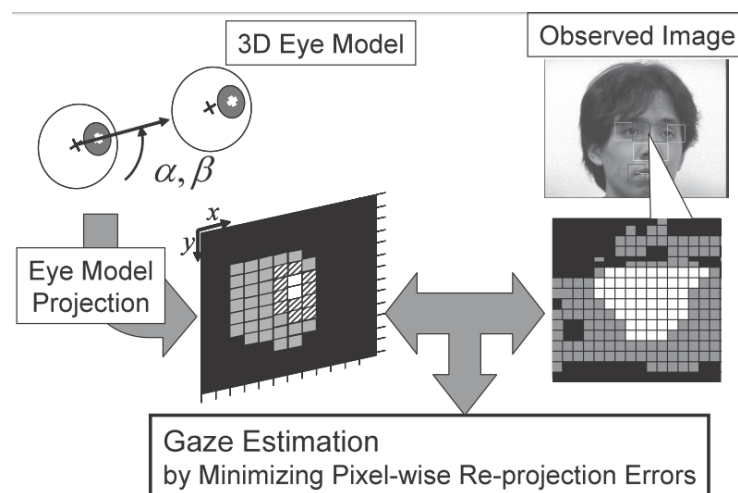


図 2.11: 眼球モデルと観測画像の誤差 [14]

Ishikawa ら [16] は顔の特徴点追跡から単眼カメラで運転手の視線を検出するシステムを提案している．運転手の視線情報から居眠り検知や，見つめている方向に危険はないかを知ら

せるなど運転支援が目的である．運転手の顔を AAM によって追跡することで図 2.12 に示す特徴点の 3 次元座標が得られる．AAM は 2 次元の特徴点追跡だが，事前に 3 次元の顔特徴点を学習させ，3D Morphable Models[17] によって 3 次元姿勢推定を行っている．この点群から 3 次元眼球モデルを用いて視線を推定できる．視線を眼球中心と虹彩中心を結んだベクトルと定義し，眼球モデルを用いて特徴点から眼球中心の値を求める．また眼球領域を切り取り，2 値化画像とエッジ画像に変換し虹彩を表した黒円画像とテンプレートマッチングを行う．それによって得られた黒円の中心を初期値とし，虹彩の輪郭に適合するように円を楕円へ変化させ，正確な虹彩中心を求める．視線の平均誤差は約 3.2 度である．実際に運転手の視線を推定する際は，車にカメラを 2 台搭載する．1 台は運転手の顔を撮影し提案システムによって視線推定を行う．もう 1 台は運転手が見ている車の外を撮影することで，運転手の注視対象が分かり，運転中に障害物に気付いているかなど計測が可能になる．

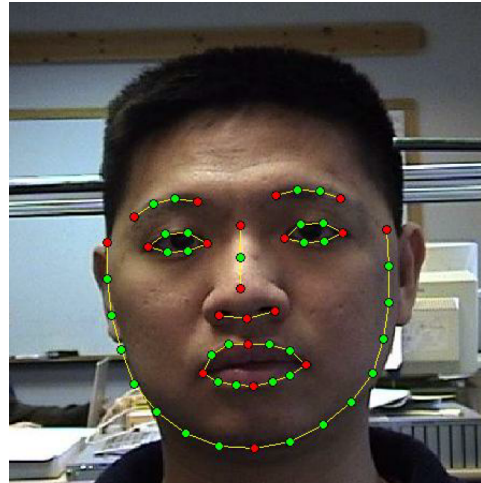


図 2.12: AAM の特徴点追跡 [16]

視線検出手法を表 2.1 にまとめる．タブレット PC で視線検出を行う際は，設置の容易さとユーザへの負担の少なさが重要となる．また，視線ジェスチャは正確な注視点を必要としない．そのため本論文では画像処理による手法によって視線検出を行う．

表 2.1: 視線検出手法の比較

視線検出手法	精度	設置の容易さ	ユーザへの負担
EOG 法			
サーチコイル法			
強膜反射法			
角膜反射法			
画像処理による手法			

## 2.2 視線を用いたインタフェース

視線インタフェースはコマンド型インタフェースと非コマンド型インタフェースに分類 [18] できる。コマンド型インタフェースは、利用者が意図的に視線を使いマウス操作 [19] などを行う方法である。視線のみを用いたコマンド型インタフェースは、人が一点を正確に見続けることが難しいため操作が困難になる問題がある。しかし高速で自然な操作を行うことができる。また、視線のみを用いて操作が行えるため、車椅子の操作 [20] など身体に障害を持つ人への支援が可能である。非コマンド型インタフェースは、視線の動きから利用者の興味や意図を推測するものである。操作を支援するだけでなく視線から利用者の興味や関心を知ることによって、適切な操作支援が行える。意図した操作を行わないため、前述した腕が使えない困難な状態での使用には向かない。しかし、閲覧者の視線から注目している情報を知ることができ、ロボットとのコミュニケーション [21] やユーザの興味から適切な商品を提示する [22] などに利用できる。本論文では意図的なページ操作が目的であるため、提案手法ではコマンド型インタフェースによる視線操作を行う。

コマンド型インタフェースには重要な問題として “Midas Touch Problem” [2] がある。これはユーザが視線操作のために意図的に見ているのか、意図せずに見ているかが区別できないことである。この問題への対処として、注視点の停留時間、または注視点の軌跡を利用する手法がある。注視点の停留時間を利用するインタフェースは、例えば選択をする際にアイコンを一定時間見続けるなど、時間的な制約で意図する操作を区別する。直感的な操作を行うことができるが、人が一点を正確に見続けることが難しく、疲れるため操作が困難になる問題がある。また、アイコンを正確に見る必要があるため、視線推定に高い精度が要求される。注視点の軌跡を利用するものは視線ジェスチャと呼ばれ、注視点の動きから意図した操作を区別する。注視点の停留時間を利用するものと比べ、マウスのように細かい操作ではなく、特定のコマンド操作しか行えない。また特定の動きをしなければならないので直感的な操作が行えない。しかし一点を見続ける必要がないため、容易で疲れにくい操作が可能となっている。

### 2.2.1 注視点の停留時間の利用

阿部ら [23] は視線による重度肢体不自由者向けのコンピュータ操作手法を提案している。図 2.13 の赤枠で囲まれた箇所を指標とし、これを注視することで操作を行う。例えば図 2.13 では文字入力操作を行っており、文字の書かれた指標を注視することでその文字の入力ができる。注視状態判定を高精度に行うため、注視状態を注視に入る前の停留状態とそれに続く継続状態に分けている。そのため、ある指標を 1 秒間注視すると停留状態となり、停留状態となった後 2 秒間は注視継続判定を行う。これは 2 秒間の間に最も注視された指標を、ユーザが注視している指標と判定している。図 2.14 は停留状態から注視継続状態への遷移を示しており、数字は指標の番号である。そのためユーザが 3 番の指標を注視していると判定している。指標と作業画面は分かれており、作業によって指標の内容が異なるコマンド操作である。そのため誤検出は少ないが入力に時間がかかる。

相田ら [24] は視線操作と他の手法によるアイコンクリックの比較を行っている。図 2.15

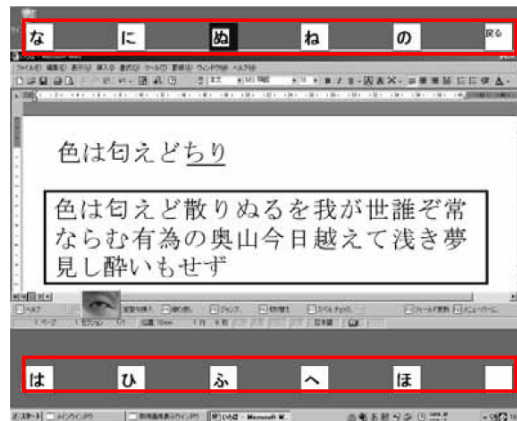


図 2.13: コンピュータ操作支援システムの指標配置 [23]

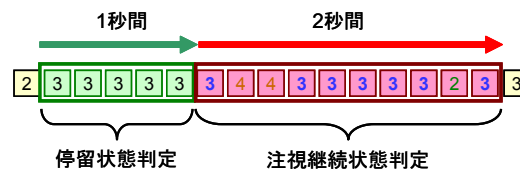


図 2.14: 注視状態判定による入力システム

に示すように水平方向または垂直方向にアイコンが並び、丸型のポインタを動かすことでアイコンの選択を行っている。マウス操作、視線のみの操作、視線とキーボードの併用操作、視線とウインクの併用操作の4種類の手法を比較した。マウス操作は通常のマウス操作である。視線のみの操作はポインタを視線で動かし、アイコンを166ms以上注視すれば選択される。視線とキーボードの併用操作はポインタを視線で動かし、キーボードのスペースキーを入力すれば選択される。視線とウインクの併用操作はポインタを視線で動かし、右目でウインクすることで選択される。その結果、視線のみの操作が最も高速な操作となった。識別率はマウス操作と比較して高くないが、直感的な操作が可能のため特に高齢者は他の手法より使いやすいと感じている。

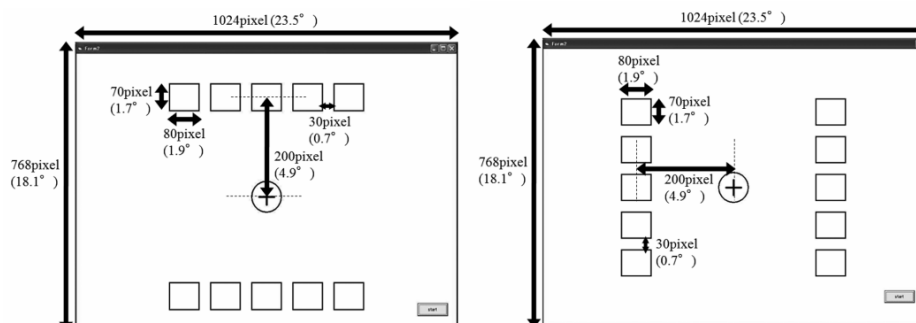


図 2.15: 実験の作業風景 [24]

### 2.2.2 注視点の軌跡の利用

Istance ら [25] は肢体不自由者のオンラインゲーム参加を目的とした、視線ジェスチャによるゲーム操作システムを提案している。視線ジェスチャを限られた時間内に行われる定義可能な眼球運動パターンと定義し、図 2.16 に示す注視対象を用いることでジェスチャを識別する。青色の部分にメジャー領域とし、ある領域から別の領域へ移ったときの注視点の軌跡を 1 本の視線ジェスチャとしている。視線ジェスチャは中央のメジャー領域から始まり、別のメジャー領域へ行き、再び中央に戻る軌跡としている。2 本と 3 本の視線ジェスチャがあり、2 本ジェスチャは 4 種類、3 本ジェスチャは 8 種類存在する。例えば、図 2.16(a) は 2 本ジェスチャであり、図 2.16(b) は 3 本ジェスチャである。Istance らは、オンライン RPG(Role Playing Game) の “World of Warcraft ®” [26] を用いて、視線ジェスチャ操作の実験を行っている。このゲームの目的は、3 次元の世界でプレイヤーの分身であるキャラクターを動かす、他のプレイヤーと協力して冒険し、モンスターと戦うことである。2 本ジェスチャは移動コマンド、3 本ジェスチャは攻撃・魔法選択、ターゲット選択が割り振られている。図 2.17 は視線ジェスチャによってゲームを操作している様子である。これらの視線ジェスチャによって経験者に実際にゲームを行ってもらったところ、ジェスチャ操作でゲームキャラクターが戦闘を行わせることが困難であることが分かった。特定の敵キャラクターへの移動やターゲット選択は難しいためである。しかし、魔法や移動は容易に行える。

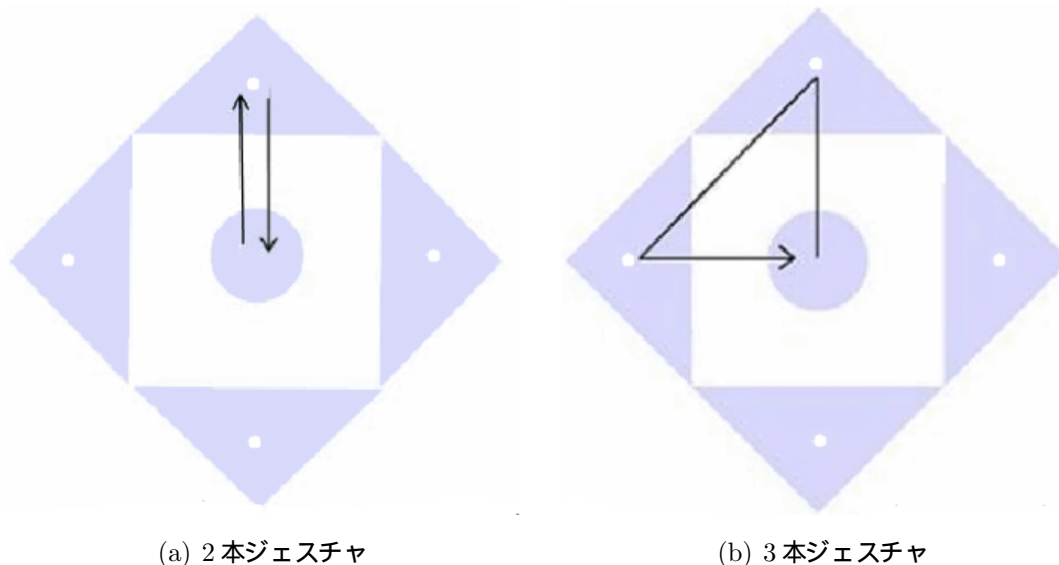


図 2.16: 注視領域と視線ジェスチャ[25]

Drewes ら [27] は視線ジェスチャによって携帯電話を操作するシステムを提案している。視線ジェスチャを相対的な眼球運動と定義し、複雑なキャリブレーションを行わずに視線操作が行える。携帯電話の四隅とそれを結ぶ 6 つの線に沿った眼球運動で操作が行える。視線ジェスチャの線を長くすることで、小さな眼球運動に反応しないようにするためである。この時、図 2.18 に示すように視線ジェスチャの方向を定義している。限られた時間内に図 2.19 に示すような眼球運動があった場合、視線ジェスチャが検出され操作が行える。例えば





図 2.17: 視線ジェスチャによるキャラクターの操作 [25]

図 2.19(a) は Web ブラウザを開く操作であり，図 2.19(b) は Web ブラウザを閉じる操作である．Drewes らは携帯電話を用いて提案手法である視線ジェスチャと従来の停留時間を用いた視線操作を比較している．視線検出手法は赤外線を用いているため，停留時間を用いた視線操作では図 2.20 に示すように携帯電話や顔を固定して実験を行っている．視線ジェスチャによる操作では，相対的な眼球運動で十分なため，携帯電話のみを固定して行っている．その結果，停留時間を用いた操作の方が直感的な操作が行えるが，視線ジェスチャの方は複雑なキャリブレーションや顔の固定を必要とせず，誤操作が少ないことが分かった．

これらの手法は角膜反射法によって注視点を検出しているため，顔がずれると視線ジェスチャが検出できない．そのため顔の固定が必要であり，携帯デバイスでの実装が困難である．本稿ではタブレット PC のみを用いた視線ジェスチャによるページ操作を目指す．

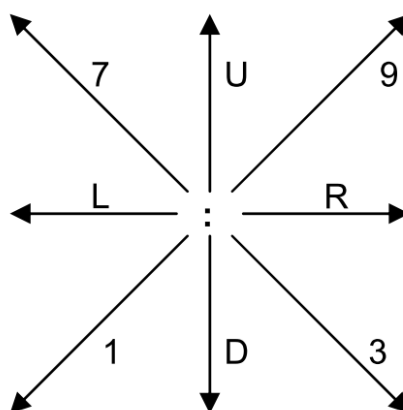


図 2.18: 視線ジェスチャの方向 [27]

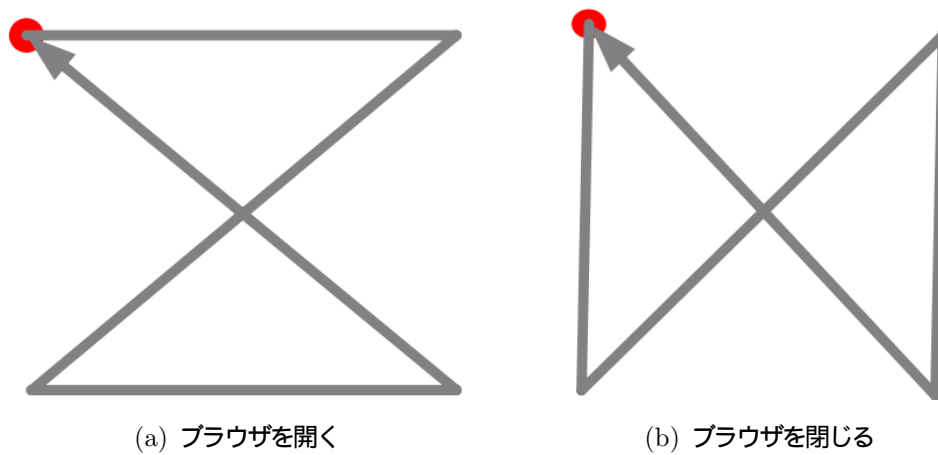


図 2.19: 視線ジェスチャによるブラウザ操作コマンド [27]

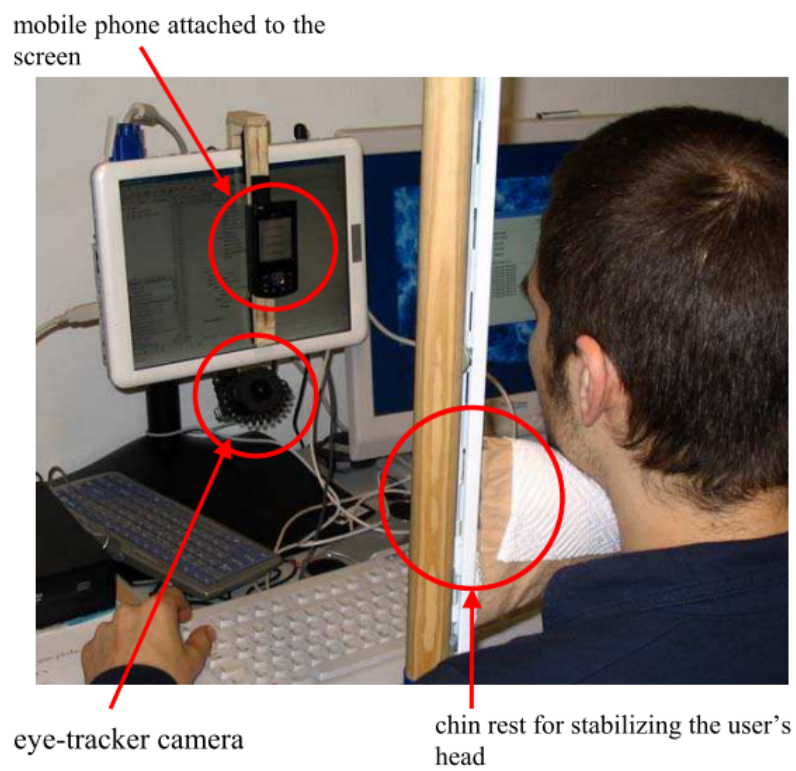


図 2.20: 実験環境 [27]

## 第3章 視軸推定

本論文ではタブレット PC のみを用いた視線ジェスチャによるページ操作を目指す．タブレット PC にはユーザの方向を向いたカメラが搭載されているので，これを用いて画像処理による手法で視線ジェスチャが計測できる視線推定を行う．提案システムの処理概要を図 3.1 に示す．図 3.1(a) は注視点の軌跡を用いた手法である．ある領域から別の領域へ移ったときの注視点の軌跡を視線ジェスチャと定義する．領域の注視が識別可能なら，注視点推定が高精度ではなくとも視線ジェスチャの識別が可能である．しかし，注視点推定にキャリブレーションが必要なため，頭部の位置が動くと正確な視線ジェスチャが得られない．そのため図 3.1(b) に示す視線ベクトルを用いた手法を提案する．視線ベクトルの大きな角度変化であるサッカード [28] によって視線ジェスチャを定義するため，頭部の位置に依存しない視線ジェスチャの識別が可能である．

### 3.1 特徴点追跡

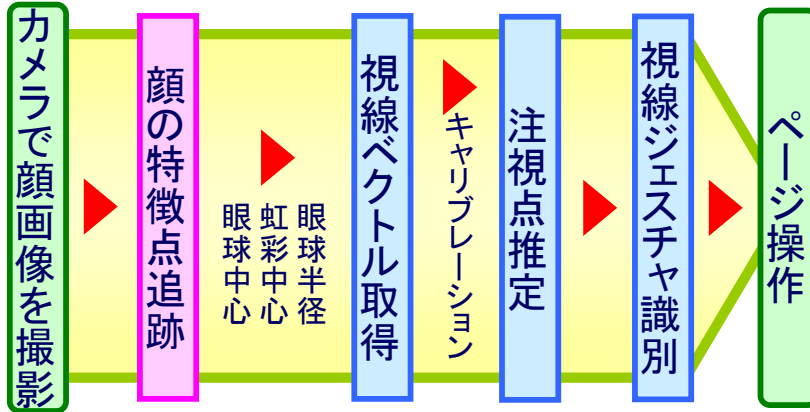
カメラでユーザの顔を撮影する．撮影した画像に対して Harr-like 特徴を用いて顔の初期位置の検出を行う．検出された顔に AAM(Active Appearance Model)[13] を用いた特徴点追跡を行う．AAM は形状の輪郭部における画素と輝度の分布を利用した画像ノイズや形状変化に強い特徴点追跡である．AAM は顔画像に特徴点を配置した大量の画像データを形状 (Shape) と外観 (Appearance) の 2 つで表現している． $u$  個の特徴点で表現される形状  $s = (x_1, y_1, x_2, y_2, \dots, x_u, y_u)^T$  を  $n$  枚の訓練画像の平均形状と形状ベクトルで表すと， $s$  は式 (3.1) で示される．

$$s = s_0 + \sum_{i=1}^n p_i s_i \quad (3.1)$$

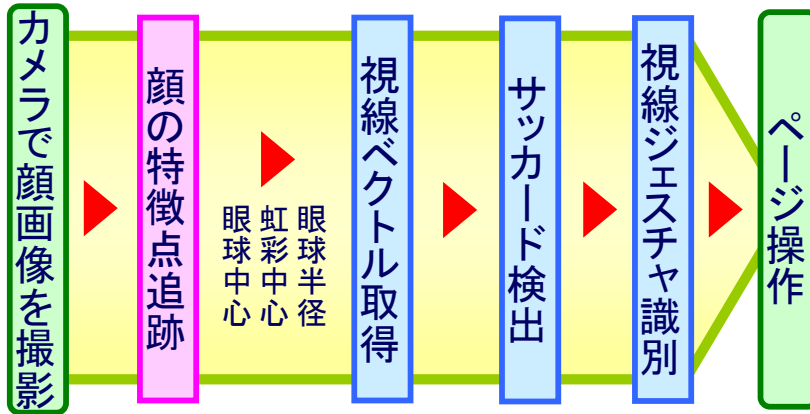
$s_0$  は  $n$  個の顔形状平均値でありこれを基本形状としている． $s_i$  は主成分分析で求めた各主成分ベクトルであり形状バリエーションを示している． $p_i$  は重み付き係数である．形状  $s$  は基本形状  $s_0$  と主成分ベクトル  $s_i$  の和であり，重み付き係数である  $p_i$  の値によって，様々な形状に変化することで，未知の顔画像に対しても柔軟な特徴点追跡が行える．同様にして，外観  $A(x)$  を特徴点  $s$  内の座標  $x = (x, y)^T$  における輝度値として考えると  $A(x)$  は式 (3.2) で示される．

$$A(x) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) \quad (3.2)$$





(a) 注視点の軌跡を用いる手法



(b) 視線ベクトルを用いる手法

図 3.1: 提案手法の処理概要

$A_0(x)$  は座標  $x$  における  $m$  個の輝度値の平均値であり  $A_i(x)$  は各主成分ベクトルである． $\lambda_i$  は重み付き係数であり，この値によって様々な外観に変化する．図 3.2 に示すように顔の向きが変化しても特徴点を頑健に追跡することが可能である．AAM には特徴点群が配置された学習データが必要である．学習データは実験時に被験者ごとに様々な向きの顔画像 9 枚を用いて作成した．このとき眼球特徴点は図 3.3 に示すように追跡される．画像上で  $i$  番目の特徴点の座標を  $(a_i, b_i)$  とする．眼球中心は眼球の特徴点から求める．眼球中心の座標を  $(u_b, v_b)$  とすると，式 (3.3) で表せる．

$$\begin{cases} u_b = \frac{a_2 + a_4}{2} \\ v_b = \frac{b_1 + b_3}{2} \end{cases} \quad (3.3)$$

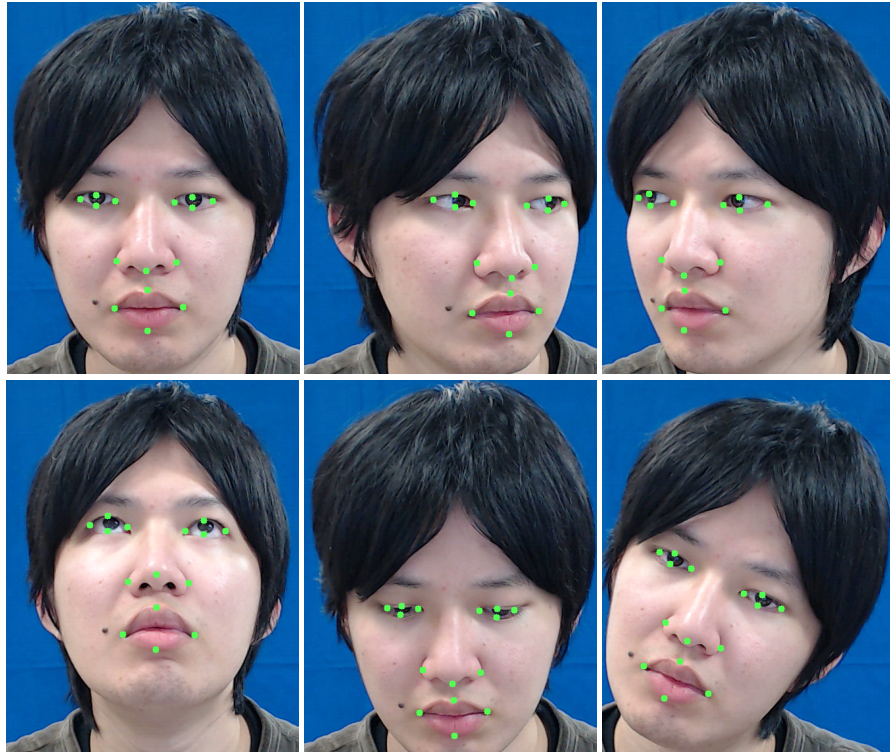


図 3.2: 顔の特徴点追跡



図 3.3: 眼球の特徴点

## 3.2 虹彩中心の検出

眼球の特徴点を用いて  $(a_1, b_2)$  と  $(a_3, b_4)$  を左上, 右下頂点とする矩形領域を眼球画像とする．眼球画像をグレースケール画像に変換し, 画像のノイズを減らすためにガウシアンフィルタによる平滑化を行い, 2 値化によって虹彩を抽出する．2 値化は P タイル法を用いる．P タイル法は, 画像におけるヒストグラムの割合による 2 値化であり, 眼球における虹彩の割合は 2 割として 2 値化を行った．その結果, 図 3.4(a) に示すように虹彩の一部が欠けた状態で抽出され, さらに一部の影がノイズとして残る．そのため, 黒色領域と白色領域のラベリングによって, 欠けた虹彩の修復を行う．まず図 3.4(b) に示すように 4 連結した黒色領域と白色領域をラベリングする．赤枠が黒色領域を囲む矩形領域であり, 青枠が白色領域を囲む矩形領域である．黒色領域の矩形領域内に包括された白色領域を黒く塗りつぶし, 残った黒

色領域のうち黒色の画素が最大の領域とその矩形領域内に包括される黒色領域を残し，小さな黒色領域を削除した．図 3.4(c) に示す赤領域は最大の黒色領域であり，黄領域は塗りつぶされる白色領域であり，青領域は最大の黒色領域に包括される黒色領域を示している．また，緑領域は削除される黒色領域である．その結果，図 3.4(d) に示すようにノイズを削除した状態で虹彩が抽出される．次に黒色領域として抽出された虹彩を楕円近似することで虹彩中心  $(u_i, v_i)$  を求める．まず黒色領域の重心を求め，図 3.4(e) に示すように，重心から放射線状に輪郭探索を行い輪郭点群を得る．得られる輪郭点群は 16 個である．図 3.4(f) に示すように，得られた輪郭点群に対して最小 2 乗法を用いて楕円近似を行った．そして近似された楕円の中心を虹彩中心  $(u_i, v_i)$  とした．

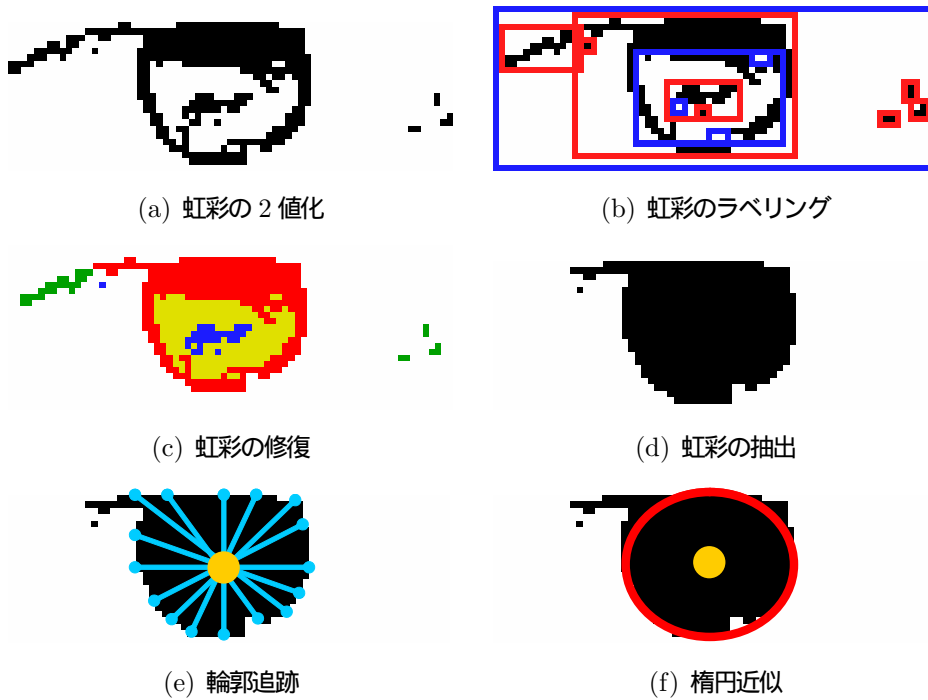


図 3.4: 虹彩抽出

### 3.3 視線ベクトル取得

眼球中心と虹彩中心，および視線ベクトルの幾何学的関係を図 3.5 に示す．眼球画像はカメラ座標系の平行投影とする．画像座標系における眼球中心  $(u_b, v_b, 0)$  と虹彩中心座標  $(u_i, v_i, 0)$  からは虹彩の偏り  $(u_i - u_b, v_i - v_b, 0)$  が求められる．眼球中心から虹彩中心までの奥行きを  $D$  とすると，視線ベクトル  $V$  は次式 (3.4) で表せる．

$$V = \begin{pmatrix} u_i - u_b \\ v_i - v_b \\ D \end{pmatrix} \quad (3.4)$$

視線ベクトル  $V$  はカメラ座標系における眼球中心から虹彩中心を結んだ線であり，視線ベクトル  $V$  と注視対象の交点がユーザが見ている注視点となる．この時，視線ベクトルの水平方向の角度  $\alpha$ ，鉛直方向の角度  $\beta$  は次式 (3.5)(3.6) で表せる．

$$\alpha = \tan^{-1} \frac{u_i - u_b}{D} \quad (3.5)$$

$$\beta = \tan^{-1} \frac{v_i - v_b}{D} \quad (3.6)$$

眼球の奥行き  $D$  は眼球画像から直接求めることはできないため，目幅  $e$  と眼球半径  $R$  を用いる．目幅と眼球半径は個人差が少なく，目幅は約  $28.4mm$  であり，眼球半径は約  $12mm$  である．この時，図 3.6 の対応関係から目幅  $e$  と眼球半径  $R$  は次式 (3.7) の関係を持つ．

$$\begin{aligned} 12 : 28.4 &= R : e \\ R &= \frac{12}{28.4} e \end{aligned} \quad (3.7)$$

目幅  $e$  は眼球特徴点の  $(a_1, b_1)$  から  $(a_3, b_3)$  までの距離なので，次式 (3.8) で表せる．

$$e = \sqrt{(a_3 - a_1)^2 + (b_3 - b_1)^2} \quad (3.8)$$

これを式 (3.7) に代入すると眼球半径  $R$  は次式 (3.9) で表せる．

$$R = \frac{12}{28.4} \sqrt{(a_3 - a_1)^2 + (b_3 - b_1)^2} \quad (3.9)$$

目幅  $e$ ，眼球半径  $R$ ，奥行き  $D$  の図 3.5 に示す関係から奥行き  $D$  は目幅  $e$ ，眼球半径  $R$  を用いて次式 (3.10) で表せる．

$$\begin{aligned} R^2 &= D^2 + (u_i - u_b)^2 + (v_i - v_b)^2 \\ D &= \sqrt{R^2 - (u_i - u_b)^2 - (v_i - v_b)^2} \end{aligned} \quad (3.10)$$

これを式 (3.5)(3.6) に代入すると水平方向の角度  $\alpha$ ，鉛直方向の角度  $\beta$  は次式 (3.11)(3.12) のようになる．

$$\alpha = \tan^{-1} \frac{u_i - u_b}{\sqrt{R^2 - (u_i - u_b)^2 - (v_i - v_b)^2}} \quad (3.11)$$

$$\beta = \tan^{-1} \frac{v_i - v_b}{\sqrt{R^2 - (u_i - u_b)^2 - (v_i - v_b)^2}} \quad (3.12)$$

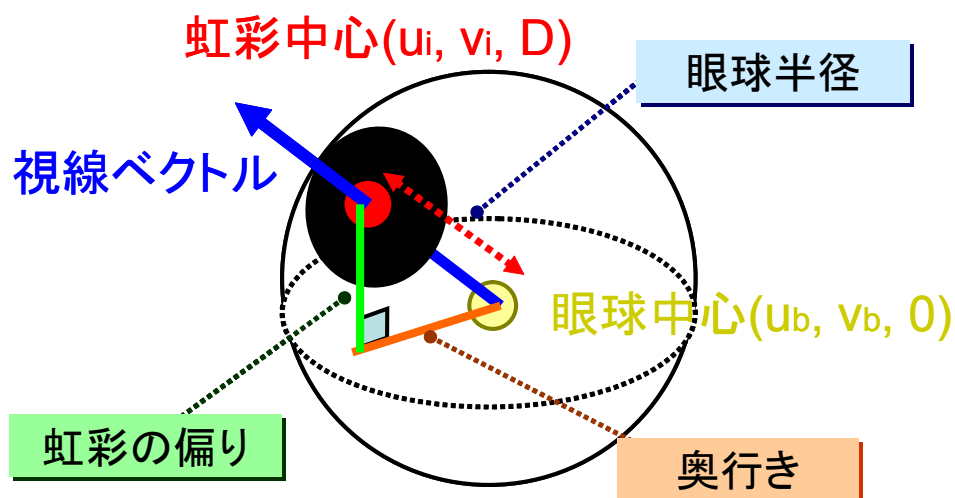


図 3.5: 眼球モデル

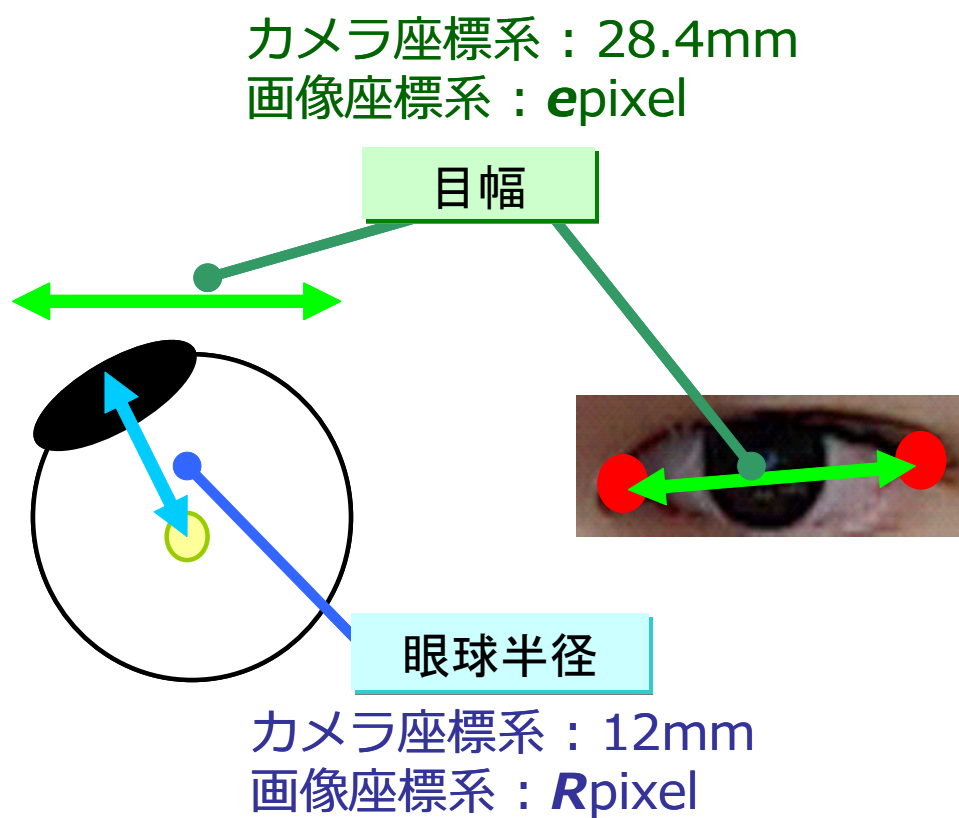


図 3.6: 目幅と眼球半径

### 3.4 注視点推定

視線ベクトル  $V$  と注視対象の交点をユーザが見ている注視点  $(p, q)$  とすると，図 3.7 の幾何学的関係から次式 (3.13) で表せる．

$$\begin{cases} p = l(\tan\alpha - \tan\Theta_{cH}) + \frac{W}{2} \\ q = l(\tan\beta - \tan\Theta_{cP}) + \frac{H}{2} \end{cases} \quad (3.13)$$

$l$  はユーザからターゲットまでの距離であり， $W$  と  $H$  はターゲットの横幅と縦幅である．また角度  $(\Theta_{cH}, \Theta_{cP})$  はターゲットの中心点  $(\frac{W}{2}, \frac{H}{2})$  を注視したときの視線ベクトルの水平方向の角度と鉛直方向の角度である． $l$ ， $(\Theta_{cH}, \Theta_{cP})$  を求めるために，水平方向と鉛直方向でキャリブレーションを行った．水平方向のキャリブレーションでは，ユーザは左端  $(\frac{W}{6}, \frac{H}{2})$ ，中央  $(\frac{W}{2}, \frac{H}{2})$ ，右端  $(\frac{5W}{6}, \frac{H}{2})$  の点をそれぞれ 2 秒間見つめる．そのとき得られた視線ベクトル角度の中央値をその点における視線ベクトルの角度  $\Theta_l$ ， $\Theta_{cH}$ ， $\Theta_r$  とする．この時  $l$  は図 3.8 より幾何学的関係から次式 (3.14) のように表せる．

$$\begin{aligned} l \tan\Theta_l + l \tan\Theta_r &= \frac{2}{3}W \\ l &= \frac{2W}{3(\tan\Theta_l + \tan\Theta_r)} \end{aligned} \quad (3.14)$$

### 3.5 評価実験

提案システムで注視点の計測実験を行った．実験ではタブレット PC の上部に装着した Web カメラを用いて被験者を撮影し，デスクトップ PC で視線の解析を行う．注視するターゲットはタブレット PC の画面に表示させる．

#### 3.5.1 被験者

本論文では 7 人の被験者に対して実験を行った．参加者は大学に通う学生で全員 20 代の男性である．顔の特徴点追跡のために，実験前に様々な方向を向いた顔画像を撮影した．撮影した顔画像は正面，上下左右，斜め方向の全 9 枚である．また被験者のうち 3 名は眼鏡によって視力を矯正している状態で顔画像の学習と視線追跡を行った．

#### 3.5.2 実験手順

図 3.9 に示すように被験者はタブレット PC から 533mm 離れた場所で顔の向きをあご台で固定して行った．この距離はでタブレット PC と頭部の距離の平均を求めることで行った．

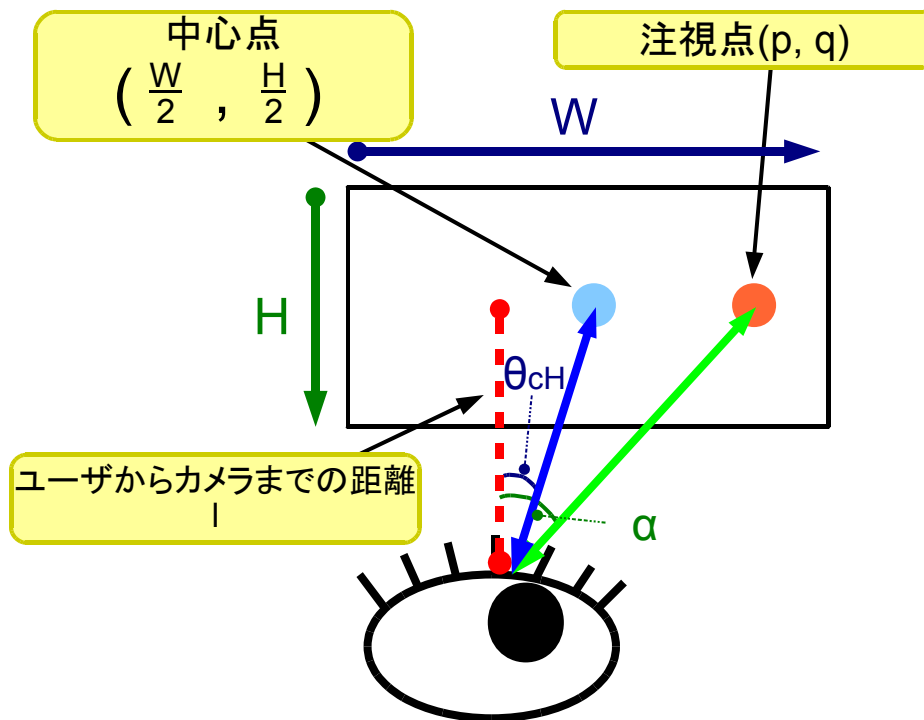


図 3.7: 注視点推定

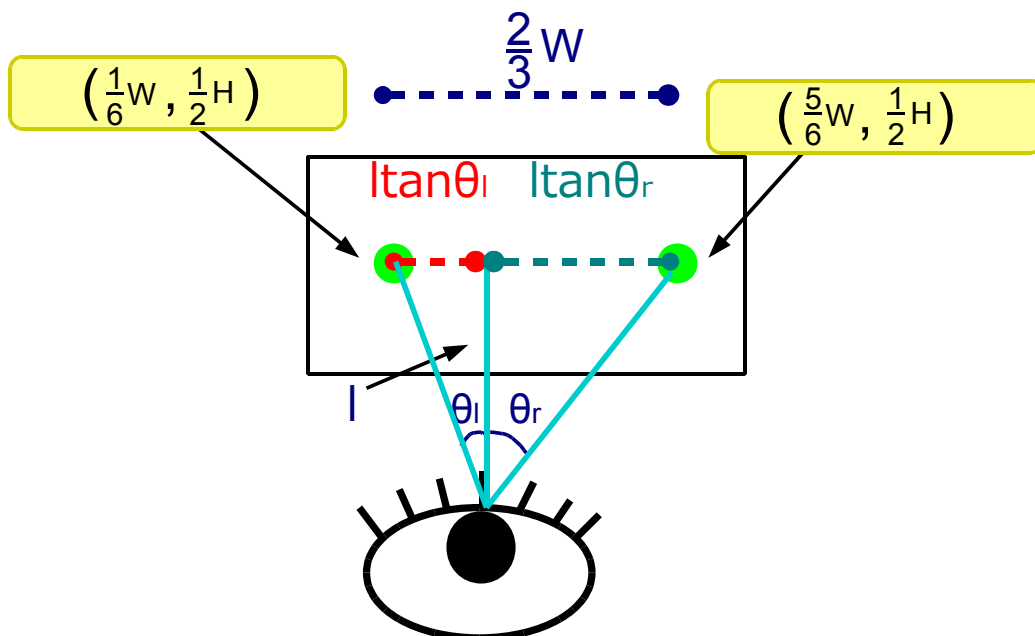


図 3.8: キャリブレーション



そして Web カメラを用いて被験者の顔を撮影し視線を実時間で計測した．実験に使用したタブレット PC は iPad® であり，使用した Web カメラは Logicoool® の “HD Pro Webcam C920t” である．撮影した画像のサイズは  $800pixels \times 600pixels$  である．タブレット PC 上に図 3.10 に示す丸い点を持つターゲットを表示し，被験者は “A1” から “C3” までの点を順に 2 秒間注視する．そのとき得られた注視点の水平方向と鉛直方向の座標を記録する．

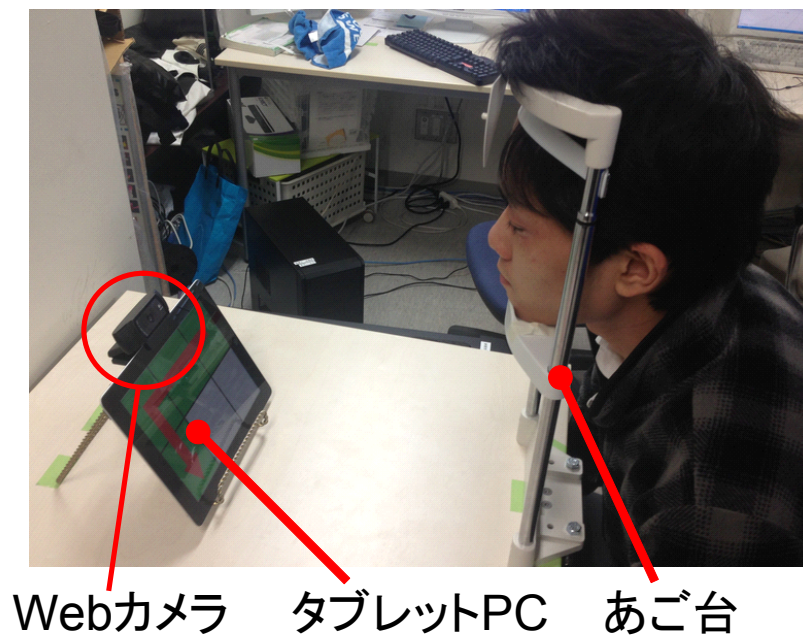


図 3.9: 実験の様子

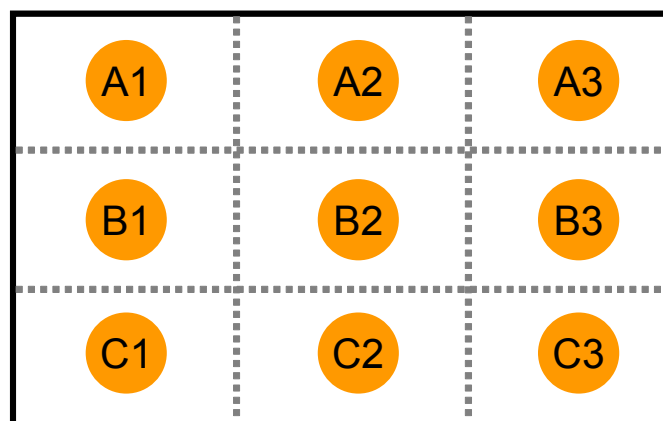


図 3.10: 注視ターゲット

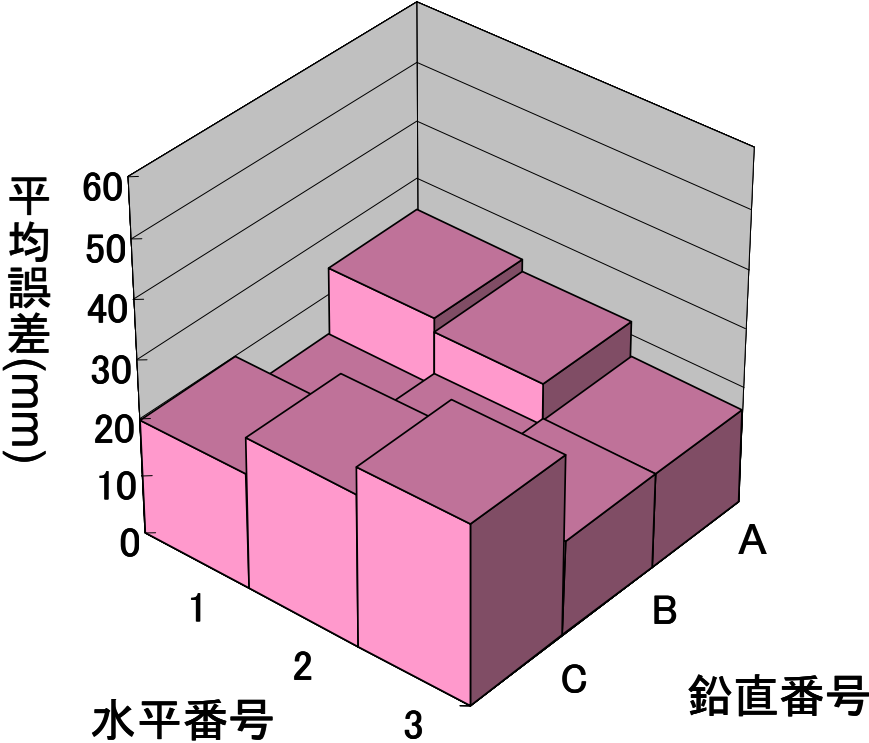


### 3.5.3 実験結果

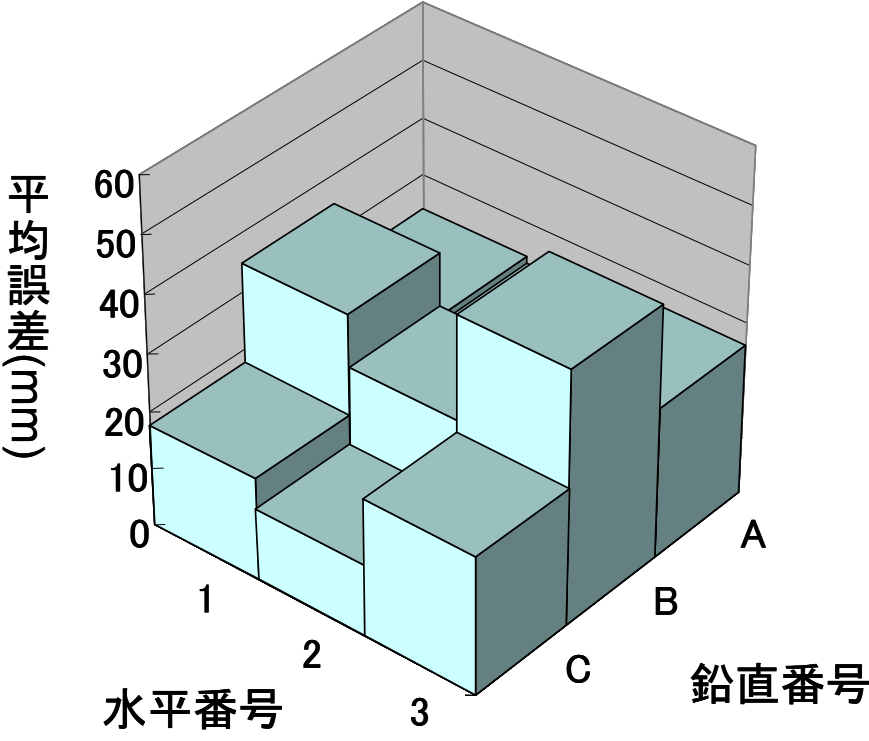
正確な座標との両目の平均誤差を，ターゲット毎にまとめたものを図 3.11 に示す．図 3.11(a) は水平方向，図 3.11(b) は鉛直方向の平均誤差である．表 3.1 は注視点の平均誤差と標準偏差である．右目と左目では大きな差はなく注視点推定が行えている．図 3.11(a) に示すように水平方向の平均誤差は安定している．特に中央を見るときは虹彩があまり隠れないため，平均誤差が小さくなっている．図 3.11(b) に示すように鉛直方向では上部  $A1 \sim A3$  や下部  $C1 \sim C3$  のターゲットを注視しているときは平均誤差が低い，中央  $B1 \sim B3$  のターゲットの平均誤差は安定しない．これは眉毛や瞼によって虹彩が隠れてしまうためである．特に左端や右端を見ているときは，虹彩の隠れる割合が大きくなるので，平均誤差も大きくなる．また，鉛直方向は水平方向と比べて平均誤差と標準偏差が大きく，正確な注視点が安定して得られていない．眼球は縦幅が小さいので鉛直方向の変化は水平方向に比べて小さいためである．

表 3.1: 注視点の平均誤差と標準偏差 ( $mm$ )

	左目		右目		両目の平均	
	水平方向	鉛直方向	水平方向	鉛直方向	水平方向	鉛直方向
平均誤差	21.7	26.8	19.1	24.7	20.4	25.7
標準偏差	17.9	18.4	15.6	20.8	16.9	19.6



(a) 水平方向



(b) 鉛直方向

図 3.11: 注視点の平均誤差



## 第4章 注視点による視線ジェスチャ

提案システムの精度から注視点の軌跡による視線ジェスチャを定義し、視線ジェスチャの評価実験を行った。

### 4.1 ジェスチャの定義

両手が塞がる作業と並行しながら携帯デバイスで書籍などを閲覧する際は、ページを送る、戻す操作が重要である。そのため、この2つのコマンドを意図する視線ジェスチャを次のように定義する。3.5.3項の結果からタブレットPCのディスプレイを図4.1に示すように $2 \times 3$ の6領域に分ける。上部の領域を“A1”，“A2”，“A3”とし、下部の領域を“B1”，“B2”，“B3”とする。1つの領域のサイズは $65.7mm \times 73.9mm$ である。水平方向において、領域の中心から別の領域までの距離は $32.9mm$ であり、注視点推定の平均誤差である $20.4mm$ より長い。そのため、ユーザが領域の中心を注視したとき、提案システムは正しい領域を検出できる。ある領域から別の領域へ移ったときの注視点の軌跡を視線ジェスチャとする。ページ送り動作の視線ジェスチャを図4.1(a)に、ページ戻しの視線ジェスチャを図4.1(b)に示す。

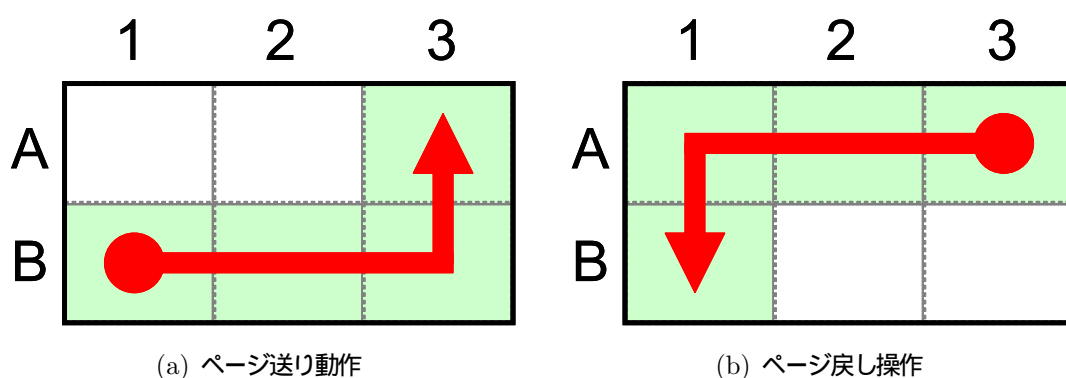


図 4.1: 注視点の軌跡を用いた視線ジェスチャによるページ操作コマンド

### 4.2 視線ジェスチャによる書籍の操作

提案システムによる視線ジェスチャ操作の評価実験を行った。実験環境は3.5.3項と同様で、被験者はそのなかから4名に対して行った。縦書きの電子書籍を図4.1に示すジェス

チャを用いて操作した．使用する注視点は右目のものである．表 4.1 はその結果である．被験者 *D* を除くと再現率が低い結果となった．これは例えば図 4.1(a) に示すジェスチャを行う場合に，眼球運動が高速であるため図 4.2 に示す領域移動が行われ，意図したジェスチャが認識されないためである．また，図 4.1(a) に示すジェスチャが失敗した場合，被験者は再びこのジェスチャを行うために視線を *B1* 領域へ移動する．この時，システムが図 4.1(b) に示す視線ジェスチャが行われたと認識してしまい，意図しない視線ジェスチャとして識別される．被験者 *D* は事前に視線ジェスチャ操作の練習を行っていたため，よい結果が得られた．

表 4.1: 注視点の軌跡による視線ジェスチャの識別精度 (%)

	A	B	C	D	平均
精度	45.0	87.5	42.9	87.5	58.0
誤識別	55.0	12.5	57.1	12.5	42.0
未識別	60.0	42.9	36.4	0	46.8

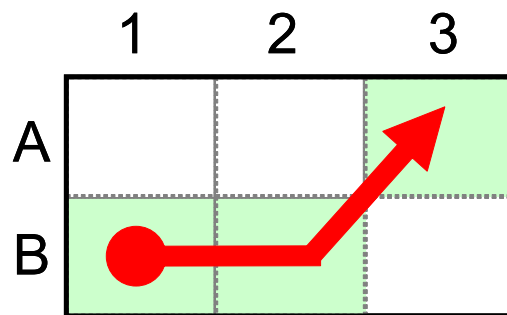


図 4.2: 高速な視線ジェスチャ

### 4.3 領域移動の遷移による視線ジェスチャの識別

注視点の移動が高速な場合は図 4.2 に示すように隣接領域を超えて軌跡が移動する．そのため，視線ジェスチャの開始領域から別の領域へ移った後，2 秒以内に視線ジェスチャの終了領域へ到達すれば，視線ジェスチャが行われたと識別する．この時，視線ジェスチャと逆向きの領域移動が行われた場合は，視線ジェスチャの識別をやり直す．また，視線ジェスチャを終えた後は終了領域から再び別の領域へ移動するまで，視線ジェスチャの識別は行わない．例えば図 4.3 に示す領域移動の遷移が行われた場合，視線ジェスチャとして識別する．しかし，図 4.4 に示す領域移動が行われた場合，視線ジェスチャとは逆向きの領域移動が行われるので，視線ジェスチャの識別をやり直す．

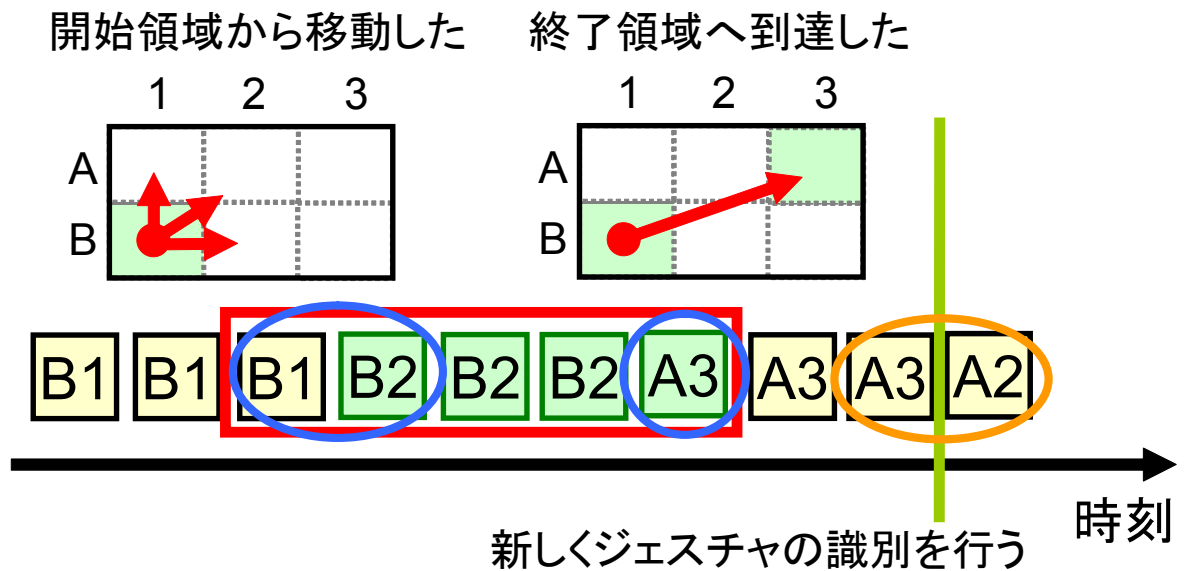


図 4.3: 領域遷移による視線ジェスチャ識別

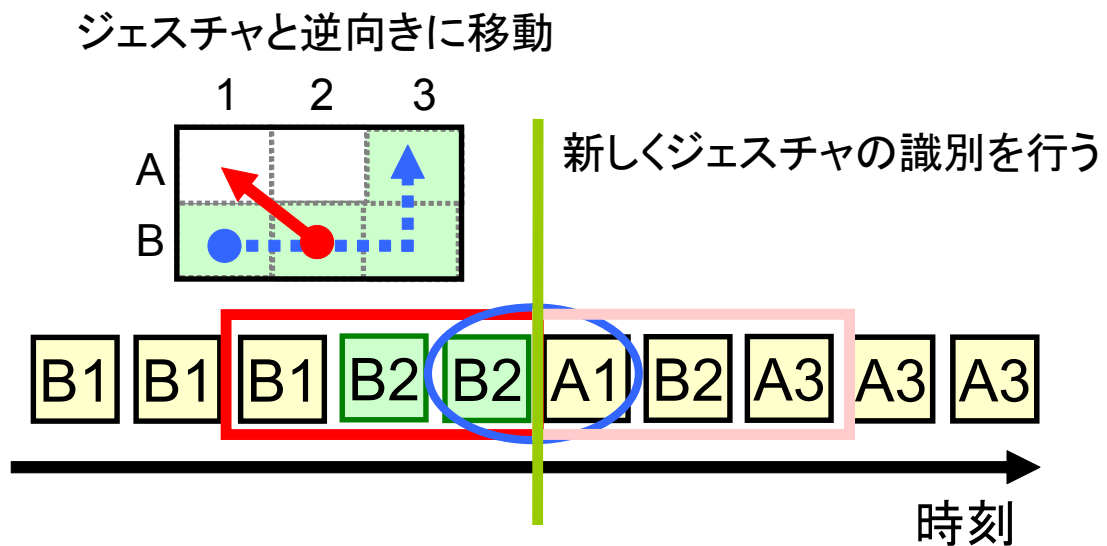


図 4.4: 視線ジェスチャ識別の失敗

## 4.4 評価実験

提案システムによる視線ジェスチャ操作の評価実験を行った。被験者と実験環境は 3.5.3 項と同様で、図 4.1 に示す視線ジェスチャがタブレット PC に 2 回ずつ表示され、表示された視線ジェスチャと同様のジェスチャを被験者が行う。視線ジェスチャの識別を行った。表 4.2 はその結果である。高速な眼球運動にも識別可能なため、識別精度が向上し、未識別は少なくなった。

表 4.2: 領域移動の遷移による視線ジェスチャの識別精度 (%)

	ページ送り動作	ページ戻し操作	平均
識別精度	71.4	60.4	66.1
誤検出	3.57	14.3	8.93
未識別	25.0	25.0	25.0

また、意図していない視線ジェスチャがどのくらいの頻度で識別されるか計測を行った。被験者は縦書きと横書きの 1 ページの文章をそれぞれ読み、その間の視線角度変化に対して提案手法が視線ジェスチャと識別した回数の平均を表 5.3 にまとめた。視線ジェスチャの向きが水平方向に大きく変化するため、横書きの場合に意図しない識別が多い。ページ送り動作は、横書きの文章の読む方向と一致しているため特に回数が多い。

表 4.3: 意図しない注視点の軌跡による視線ジェスチャ(回)

	ページ送り動作	ページ戻し動作	平均
縦書き	5.29	3.71	4.50
横書き	11.0	8.29	9.64

## 第5章 視線ベクトルによる視線ジェスチャ

注視点を用いた視線ジェスチャはキャリブレーションが必要であり，頭部が移動すると正確な注視点が得られず，視線ジェスチャが識別できない．そのため，視線角度変化による視線ジェスチャを定義する．視線角度の変化はターゲットとユーザの距離によって異なるため，タブレット PC を利用する場合の，タブレット PC からユーザ頭部までの平均距離を事前に測定した．

### 5.1 タブレット PC とユーザ頭部の距離測定

タブレット PC の画面の端から端を見るとき，視線ベクトルの変化はターゲットとユーザの距離によって異なる．そのため，タブレット PC を利用するときのタブレット PC からユーザ頭部までの距離を測定した．測定は 8 人の被験者に対して行った．実験に使用したタブレット PC は iPad® であり，Kinect® を用いて距離の測定を行った．図 5.1 に示すように被験者を見下ろすように Kinect® を設置することで，タブレット PC に邪魔されずに骨格追跡が行える．図 5.2 に示すように上半身の骨格追跡が行われ，被験者の関節毎の座標点が計測される．タブレット PC は手によって支えられるため，両手の中点をタブレット PC の位置とし，中点から頭部の座標点までの距離をタブレット PC からユーザ頭部への距離とした．机の上にタブレットを置き，ユーザは椅子に座りながら指示されたタブレット PC の操作を行う．操作として，まずインターネットの Youtube® で好きなビデオを検索し閲覧し，次に e-Book を開き好きな本を読むことを指示した．操作が終わるまで記録した距離の平均をその人物の頭部からタブレット PC への距離とする．測定の結果と平均距離，標準偏差を表 5.1 に示す．標準偏差が低いことから操作中に顔が大きく動いていないと言える．そのため，タブレット PC からユーザ頭部への距離を  $533mm$  として定める．

表 5.1: 頭部への距離の平均誤差と標準偏差 ( $mm$ )

被験者	A	B	C	D	E	F	G	H	平均
頭部への距離	562	521	605	522	441	603	535	476	533
標準偏差	30	13	29	48	45	21	28	45	53



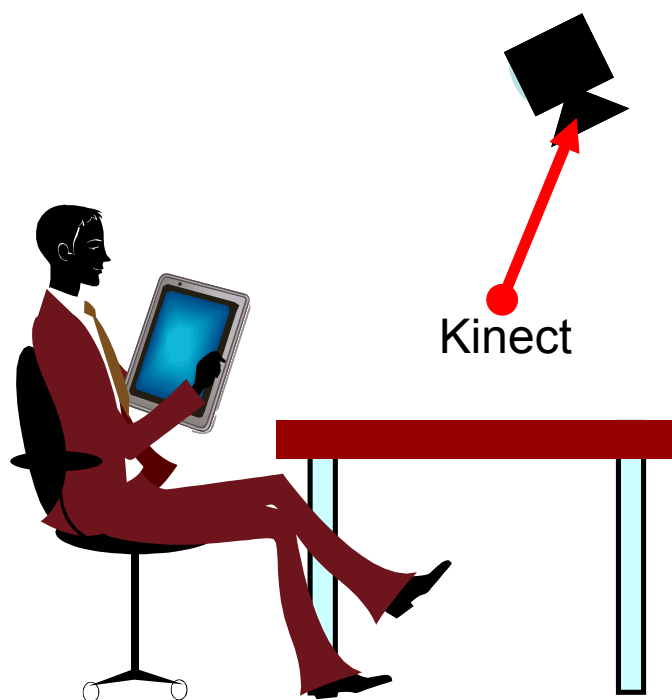


図 5.1: Kinect®による測定の様子



図 5.2: 骨格追跡

## 5.2 視線ジェスチャの定義

図 5.3 に示すようにタブレット PC の周辺に沿った眼球運動を視線ジェスチャとする．実験に使用したタブレット PC は iPad® でありモニタのサイズは  $197.1mm \times 147.8mm$  である．図 5.4 に示すように視線ジェスチャの開始位置の視線角度を  $0^\circ$  とすると，水平方向において左端領域の中心から右端領域の中心までの距離が  $131.4mm$ ，頭部からタブレット PC までの距離が  $533mm$  なので視線ジェスチャによる視線角度の変化は水平方向は  $13.9^\circ$  と定義する．同様にして下領域の中心から上領域の中心までの距離が  $98.1mm$  なので鉛直方向の視線角度の変化を  $7.94^\circ$  と定義する．両手が塞がる作業と並行しながら携帯デバイスで書籍などを閲覧する際は，ページを送る，戻す操作が重要である．ページ送り動作の視線ジェスチャを図 5.3(a) に，ページ戻し動作の視線ジェスチャを図 5.3(b) に示す．例えば図 5.3(a) に示す視線ジェスチャが行われる際，まず左端から右端へ水平方向の視線角度が大きく変化するサッカーカード [28] が行われる．その後，右端を注視した後，右上へ鉛直方向のサッカーカードが行われる．また，水平方向のサッカーカード運動が行われる際は，鉛直方向の視線角度変化は少なく，鉛直方向のサッカーカード運動が行われる際は，水平方向の視線角度変化が少なくなっている．そのため，水平方向のサッカーカード運動を検出した後，鉛直方向のサッカーカード運動を検出することで視線ジェスチャの識別が行える．提案システムの精度は角度にすると水平方向が  $2.21^\circ$ ，鉛直方向が  $2.77^\circ$  であり，特に鉛直方向は視線角度変化の値が小さいので誤識別が起こりやすい．また，サッカーカードと定義できる急激な視線角度変化は 1 フレーム間の変化では検出が困難である．そのため，数フレーム間の視線角度の変化から急激な変化をサッカーカードとして検出する．本論文では，フレーム間の角度変化の差分を用いる手法と最小 2 乗法による手法によってサッカーカードを検出し，サッカーカードの組み合わせから視線ジェスチャの識別を行った．

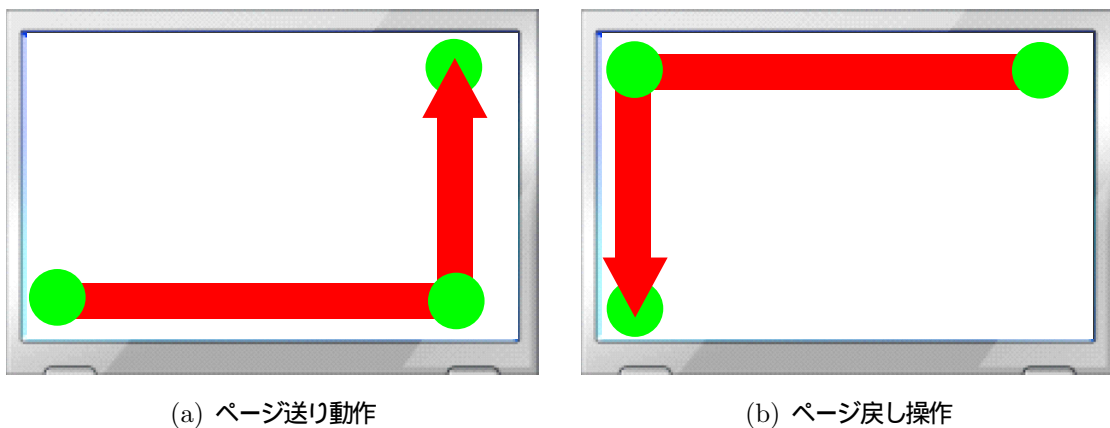


図 5.3: 視線角度変化を用いた視線ジェスチャによるページ操作コマンド

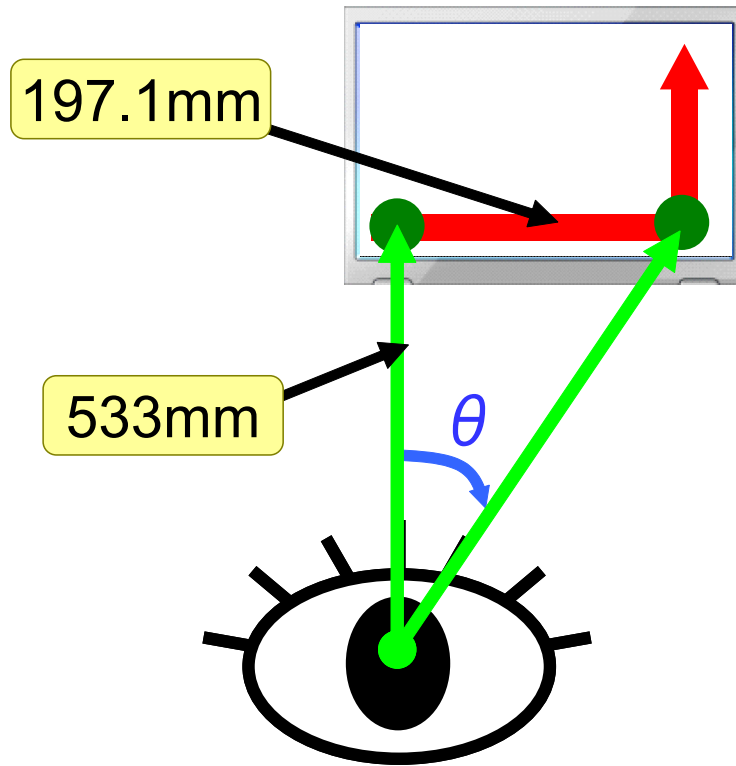


図 5.4: 水平方向のサッカード

### 5.2.1 フレーム間差分によるサッカード検出

あるフレーム  $f$  における水平方向の視線角度を  $\alpha_f$  とすると，図 5.5 に示すように条件式 (5.1) を満たすとき水平方向の正のサッカードが検出される．

$$\alpha_f - \alpha_{f-i} \geq 13.9 \quad (i = 1 \sim n) \quad (5.1)$$

同様にしてあるフレーム  $f$  における鉛直方向の視線角度を  $\beta_f$  とすると，条件式 (5.2) を満たすとき鉛直方向の正のサッカードが検出される．

$$\beta_f - \beta_{f-i} \geq 7.94 \quad (i = 1 \sim n) \quad (5.2)$$

$n$  はフレーム  $f$  から 0.5 秒前までのフレーム数である． $1 \sim n$  の間で複数のサッカードが検出された場合，フレーム  $f$  に最も近いサッカードをそのフレームにおけるサッカードとする．例えば図 5.6 に示すように視線角度が変化した場合， $\alpha_f - \alpha_{f-1}$  では負の向きのサッカード， $\alpha_f - \alpha_{f-2}$  では正のサッカードが検出されている．この時，フレーム  $f$  に最も近いのは負のサッカードである  $\alpha_f - \alpha_{f-1}$  のため，フレーム  $f$  は負のサッカードとして検出される．

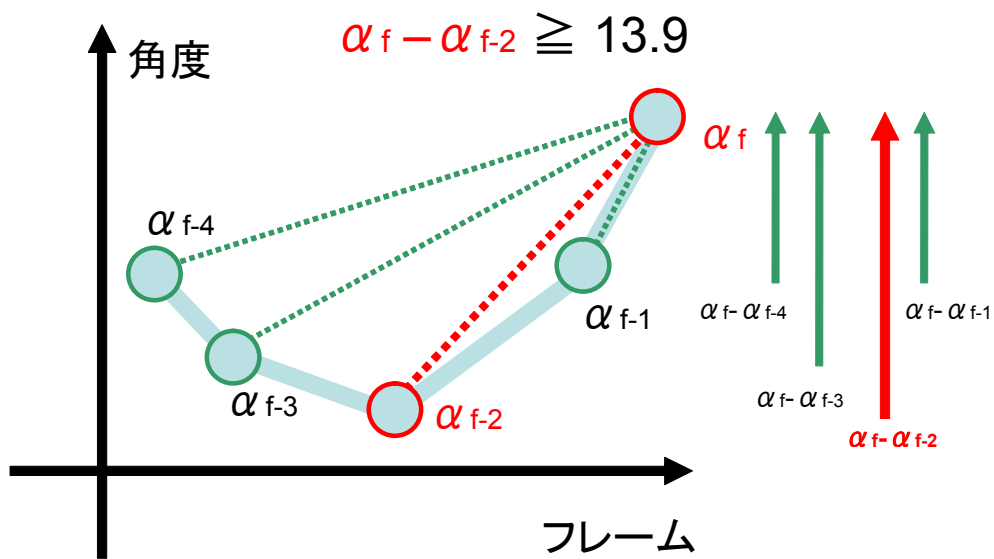


図 5.5: フレーム間差分によるサッカード検出

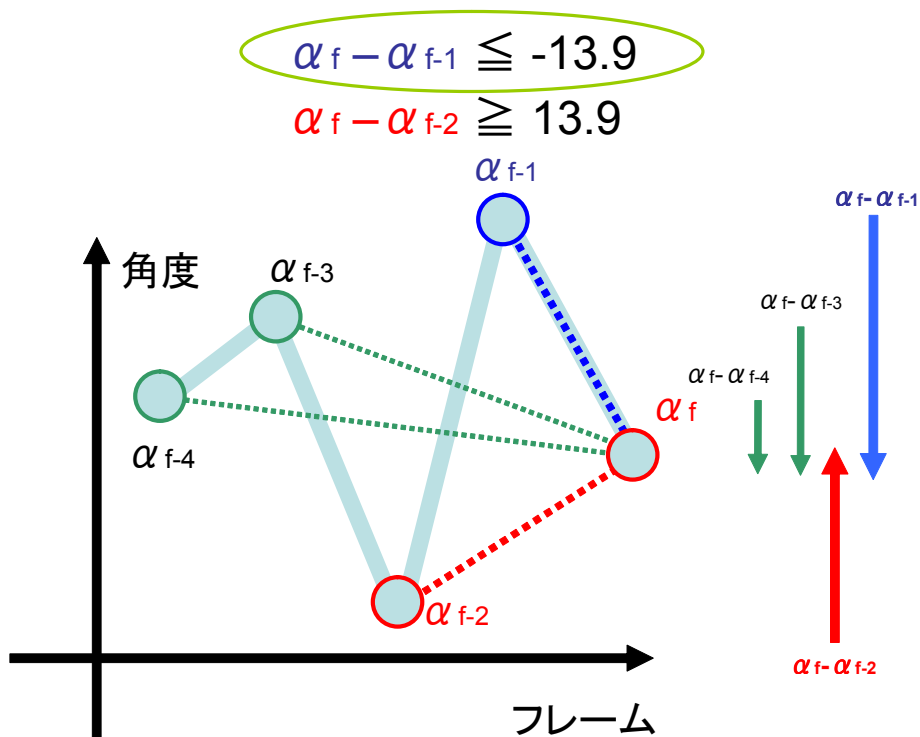


図 5.6: 優先されるサッカード

### 5.2.2 最小2乗法によるサッカード検出

サッカードによる視線角度変化は短時間の間で行われるため、数フレーム間の変化量を最小2乗法で求めることで、ノイズに強いサッカードの検出を行った。フレーム  $f$  における水平方向の視線角度の時間変化を  $(\alpha_f, t_f)$  とする。 $\alpha_f$  は  $f$  フレームにおける水平方向の視線角度の値であり、 $t_f$  は  $f$  フレームにおける時刻であり単位は秒である。あるフレーム  $f$  とフレーム  $f$  から 0.5 秒前までの  $n$  フレームの視線角度の時間変化を用いて、最小2乗法によって直線の方程式を近似する。このときの直線の傾きを  $\alpha_{Mf}$  とし、図 5.7 に示すように次の条件式 (5.3) を満たすとき水平方向の正のサッカードが検出される。

$$\alpha_{Mf} \geq 13.9 \quad (5.3)$$

同様にしてフレーム  $f$  における鉛直方向の視線角度の時間変化  $(\beta_f, t_f)$  から最小2乗法で求めた直線の傾きを  $\beta_{Mf}$  とすると、次の条件式 (5.4) を満たすとき鉛直方向の正のサッカードが検出される。

$$\beta_{Mf} \geq 7.94 \quad (5.4)$$

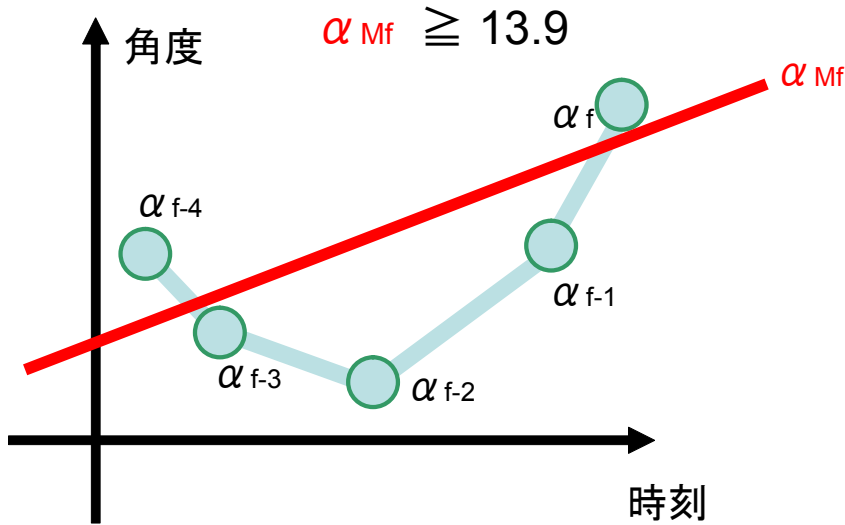


図 5.7: 最小2乗法によるサッカード検出

## 5.3 視線ジェスチャの識別

フレーム間の角度変化の差分を用いる手法と最小2乗法による手法によってフレーム毎に水平方向と鉛直方向のサッカードの検出を行う。図 5.3 で定義した視線ジェスチャでは水平方向のサッカード運動が行われる場合は、鉛直方向の視線角度変化が少なく、鉛直方向の

サッカード運動が行われる場合は、水平方向の視線角度変化が少なくなっている．そのため、サッカードの判定を次のように行う．

1. 水平方向のサッカードのみが行われた：水平方向のサッカードと判定
2. 鉛直方向のサッカードのみが行われた：鉛直方向のサッカードと判定
3. 水平方向と鉛直方向のサッカードが同時に行われた：水平方向のサッカードと判定
4. 水平方向と鉛直方向のサッカードがどちらも行われぬ：微小な視線角度変化

鉛直方向の注視点推定精度は低く、サッカードによる変化も水平方向より小さいため、意図しない鉛直方向のサッカードが頻繁に検出される．そのため水平方向と鉛直方向のサッカードが同時に行われた場合は、水平方向のサッカードを優先する．別方向のサッカードを連続で行うのは困難であため、図 5.8 に示すように水平方向のサッカードが行われ、右端を注視した後、鉛直方向のサッカードが行われる．サッカードの間に注視が行われるため、水平方向のサッカードと鉛直方向のサッカードが連続して検出されるとは限らない．そこで図 5.9 に示すように水平方向のサッカードが行われた後、1.0 秒以内に鉛直方向のサッカードが行われれば、視線ジェスチャが行われたと識別する．同じサッカードが連続で検出された場合、最後に検出された水平方向のサッカードから 1.0 秒間、鉛直方向のサッカードを判定する．フレーム毎にサッカードを検出しているため、同じサッカードが連続で起こりやすく、視線ジェスチャも連続で識別される可能性がある．そのため、図 5.10 に示すように視線ジェスチャが識別された場合、それまでのフレームを用いずに新しくサッカードの識別を行う．また、水平方向の正のサッカードと負のサッカードは独立して判定を行う．

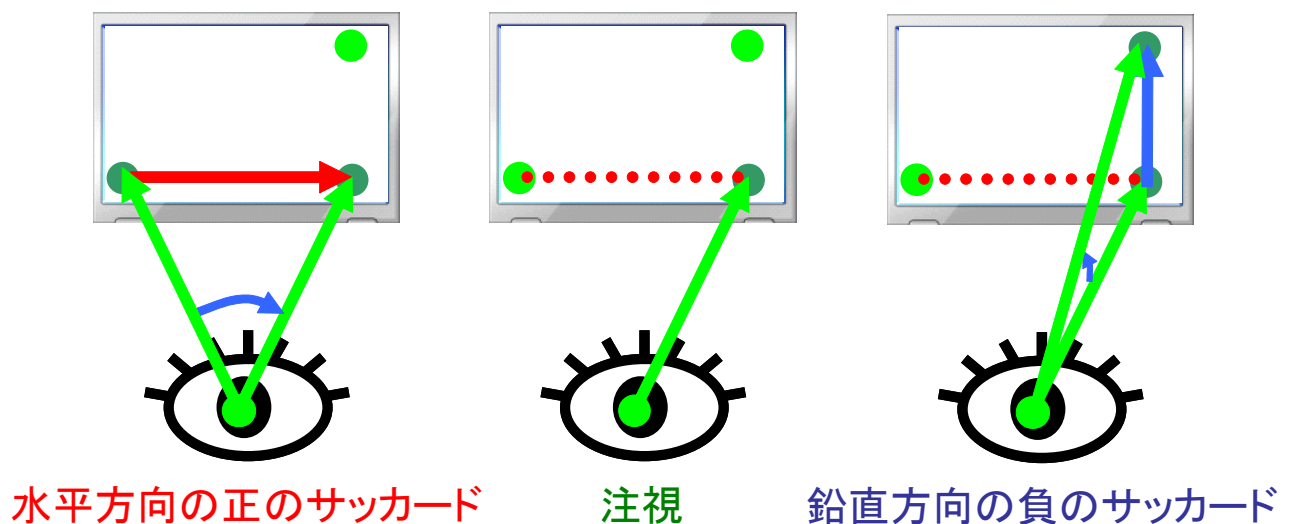


図 5.8: 視線ジェスチャの遷移

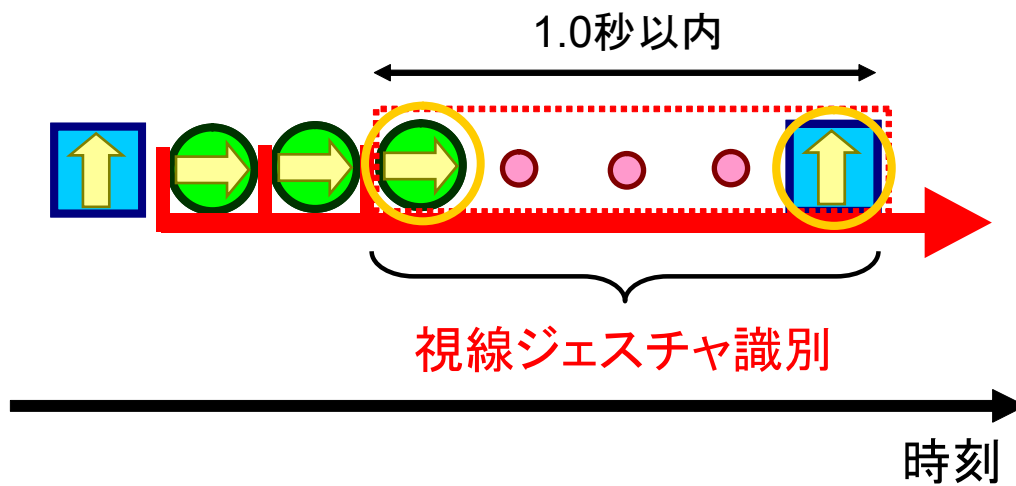
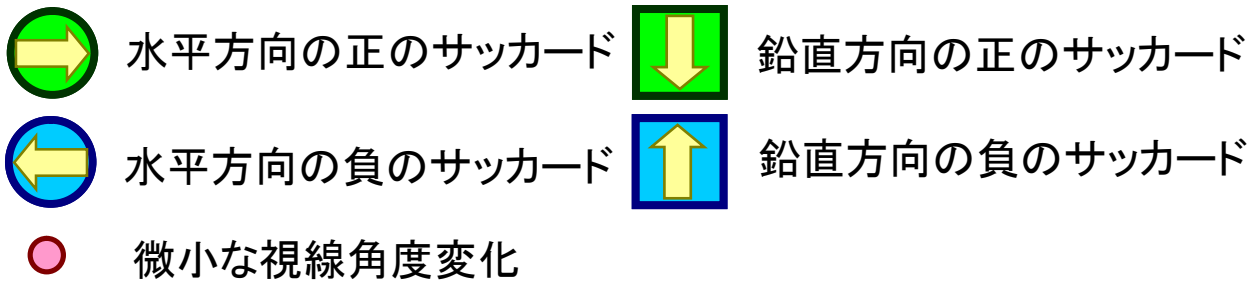


図 5.9: サッカードの遷移による視線ジェスチャの識別

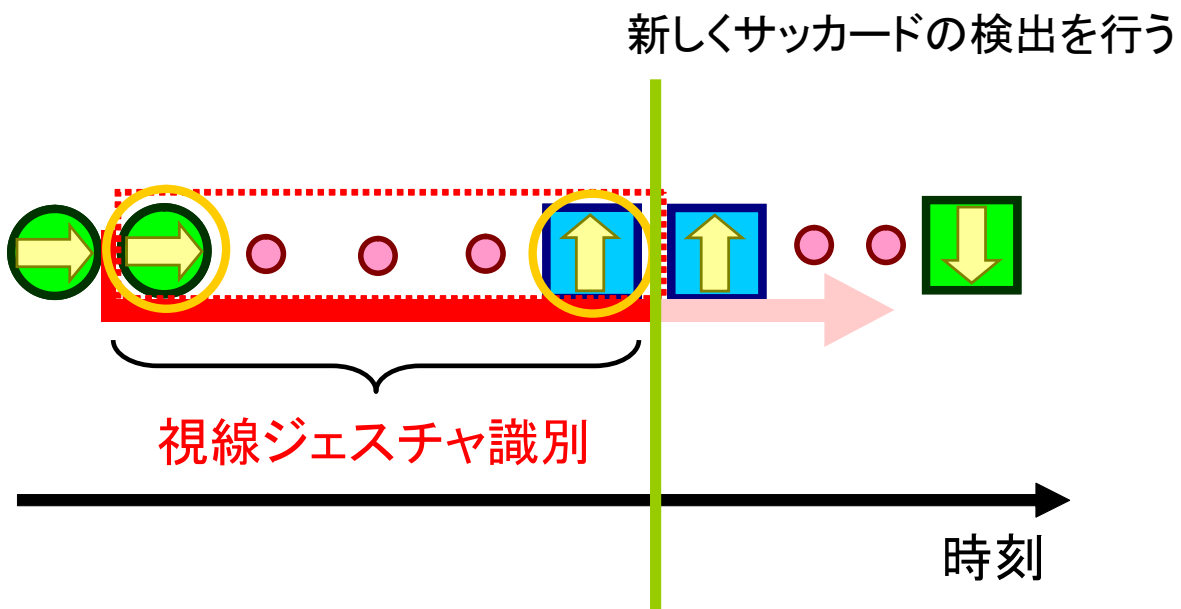


図 5.10: ジェスチャ識別後

## 5.4 評価実験

提案システムで視線ジェスチャの識別実験を行った．被験者と実験環境は 3.5.3 項と同様で，図 5.3 に示す視線ジェスチャがタブレット PC に 2 回ずつ表示され，表示された視線ジェスチャと同様のジェスチャを被験者が行う．表 5.2 はその結果である．フレーム間差分による手法は表 4.2 に示す注視点の手法より精度が低い結果となった．これはフレーム間差分による手法が誤差の影響が多きいためである．最小 2 乗法による手法は精度がより向上し，誤識別と未識別も低い結果となった．これは最小 2 乗法によって誤差の影響が少なくなったためである．そのため，最小 2 乗法による手法は表 4.2 に示す注視点の手法より高精度の視線ジェスチャの識別が可能である．

表 5.2: 視線角度変化による視線ジェスチャの識別精度 (%)

	差分			最小 2 乗法		
	ページ送り動作	ページ戻し動作	平均	ページ送り動作	ページ戻し動作	平均
識別精度	50.0	42.9	46.3	75.0	71.4	73.2
誤識別	7.14	25.0	16.1	7.14	7.14	7.14
未識別	42.9	32.1	37.5	17.9	18.6	18.3

また，意図していない視線ジェスチャがどのくらいの頻度で識別されるか計測を行った．被験者は縦書きと横書きの 1 ページの文章をそれぞれ読み，その間の視線角度変化に対して提案手法が視線ジェスチャと識別した回数の平均を表 5.3 にまとめた．最小 2 乗法による手法は表 4.3 に示す注視点の軌跡による手法と比較して，意図しないジェスチャの回数が多くなっている．これはサッカードが行われれば，注視する位置に関係なく，視線ジェスチャが識別できるためである．意図しない視線ジェスチャの回数は縦書きと横書きの違いによる影響があまり無い．これは視線角度の手法が注視点の軌跡による手法と比べて時間的制約が大きいためだと思われる．

表 5.3: 意図しない視線角度変化による視線ジェスチャ(回)

	差分			最小 2 乗法		
	ページ送り動作	ページ戻し動作	平均	ページ送り動作	ページ戻し動作	平均
縦書き	9.71	4.71	7.21	17.0	11.1	14.1
横書き	6.71	8.57	7.64	10.6	14.4	12.7





## 第6章 結論

### 6.1 本論文の結論

本論文では、タブレット PC のみを用いた視線ジェスチャによるページ操作支援を検討した。まず、タブレット PC にはユーザの方向を向いたカメラが搭載されているので、これを用いて画像処理による手法で視線ジェスチャが計測できる視線推定を行った。特徴点追跡で頑健に眼球領域を取得し、欠落した虹彩を修復することで、視線ベクトルの値を正確に取得し、視線ジェスチャの識別が可能な計測精度を得られた。

注視点の平均誤差から、視線ジェスチャが識別可能な領域に分け、注視点の軌跡による視線ジェスチャを定義した。注視点による視線ジェスチャは、ある領域から別の領域へ移った注視点の軌跡とした。注視点の誤差の影響と高速な眼球運動を考慮しなかったため識別精度は低く、未識別も高くなった。そのため、領域移動の遷移を用いて視線ジェスチャの識別を行った。これにより識別精度が向上し、未識別も少なくなった。

注視点を用いた視線ジェスチャはキャリブレーションが必要であり、頭部が移動すると正確な注視点が得られず、視線ジェスチャが識別できない問題がある。そのため、視線ベクトルによる視線ジェスチャを提案した。水平方向と鉛直方向のサッカーの組み合わせで視線ジェスチャが識別できる。サッカーの検出はフレーム間差分による手法と最小 2 乗法による手法を用いた。フレーム間差分による手法は誤差の影響が大きいので視線ジェスチャの識別精度が低く、未識別も高くなった。最小 2 乗法による手法による手法は識別精度が高く、誤識別と未識別は低くなり、注視点の軌跡を用いた手法よりも高精度の視線ジェスチャ識別が可能である。そのため、意図的な視線操作が視線ベクトルのみで行える。しかし、読書時の意図しない視線ジェスチャの回数は他の手法と比べて多かった。

### 6.2 今後の課題

本論文では、視線角度変化による視線ジェスチャの識別手法を提案した。今後の課題は視線ジェスチャの識別の向上と意図しない視線ジェスチャへの対処である。例えば視線ジェスチャの動きは、横書きの文章を読むときの眼球運動に近いので、読書時に意図しない視線の動きが視線ジェスチャとして識別されることがある。そのため、より適切な視線ジェスチャの動きを検討する必要がある。現在は決められた時間内に起きるサッカーを検出しているため、より高速なサッカーが検出されにくくなっている。そのため、より柔軟な時間的制約が必要である。また、現在は頭部を固定した状態での精度評価に留まっているが、頭部を固定せずに行った視線ジェスチャ識別の精度評価を行う予定である。

提案システムは AAM による特徴点追跡に計算時間がかかるため、デスクトップ PC で解析を行っている。しかし、視線ジェスチャの識別には注視点の高い精度は必要ないため、精度が低くとも特徴点追跡が高速な手法を用いることで、タブレット PC での実装が可能となる。また、タブレット PC を固定せずに操作する場合、顔全体が撮影されない可能性があるため、顔の一部分に対して特徴点追跡を行う必要がある。

## 謝辞

本論文の作成，研究を進めるにあたり御指導くださった高橋裕樹准教授に心から御礼申し上げます．

最後に，日頃より参考となる御意見を下さり，実験を手伝って下さいました研究室の皆様に御礼申し上げます．



## 参考文献

- [1] 大野健彦：“視線インタフェースから視線コミュニケーションへ-視線のある環境を目指して-”，情処学報 2001-HI-93, pp.47-54, 2001.
- [2] R.J.K.Jacob：“What You Look At is What You Get: Eye Movement-Based Interaction Techniques”，CHI, ACM, pp.11-18, 1990 .
- [3] 大野健彦，武川直樹，吉川厚：“2点補正キャリブレーションを実現した視線測定システム”，情処学論, Vol.44, No.4, pp.1136-1149, 2003.
- [4] H.Drewes and A.Schmidt：“Interacting with the Computer using Gaze Gestures”，Proceeding INTERACT 2007 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction, Vol.2, pp.475-488, 2007.
- [5] “Design Office アイリス・アイリス”  
<http://web.kyoto-inet.or.jp/people/irisiris/index.html>
- [6] 久野克章，八木透，藤井一幸，古賀一男，内川嘉樹：“EOG を用いた視線入力インターフェイスの開発”，情処学論, Vol.39, No.5, pp.1455-1462, 1998.
- [7] K.R.Boff, L.Kaufman and J.P.Thomas：“HANDBOOK OF PERCEPTION AND HUMAN PERFORMANCE”，A Wiley-Interscience Publication, Chapter.42, p.43, 1986.
- [8] 野田明子，古池保雄：“終夜睡眠ポリグラフィ”，生体医工学, Vol.46, No.2, pp.134-143, 2008.
- [9] 竹森節子，森山春子：“サーチコイルによる眼球運動の記録・分析”，Equilibrium Res, Vol.45, No.4, pp.310-317, 1986.
- [10] 石田正幸，川崎匡，渡辺行雄：“ネコの水平・垂直性視運動性眼振と視運動性後眼振の解析”，日本耳鼻咽喉科学會會報, Vol.107, No.3, pp.(107)179-187, 2004.
- [11] 阿部清彦，大井尚一，大山実：“強膜反射法を用いた画像解析による視線入力システム—目頭抽出による頭部移動の検出と補正—”，信学技報, Vol.102, No.418, pp.47-52, 2002.

- [12] 佐竹純二, 小林亮博, 平山高嗣, 川島宏彰, 松山隆司: “高解像度撮影における実時間視線推定の高精度化”, 信学技報, PRMU2007-238, Vol.107, No.491, pp.137-142, 2008.
- [13] I.Matthews and S.Baker: “Active Appearance Models Revisited”, International Journal of Computer vision, Vol.60(2), pp.135-164, 2004.
- [14] 山添大丈, 内海章, 米澤朋子, 安部伸治: “単眼カメラを用いた視線推定のための三次元眼球モデルの自動キャリブレーション”, 信学論 (D), Vol.J94-D, No.6, pp.998-1006, 2011.
- [15] B.D.Lucas and T.Kanade, “An iterative image registration technique with an application to stereo vision”, Proc.Int’l Joint Conf.Artificial Intelligence, pp.674-679, 1981.
- [16] T.Ishikawam, S.Baker, I.Matthews, and T.Kanade: “Passive driver gaze tracking with active appearance models”, Tech.Rep.CMU-RI-TR-04-08, The Robotics Institute, Carnegie Mellon University, 2004.
- [17] V.Blanz and T.Vetter: “A morphable model for the synthesis of 3D faces”, SIGGRAPH, pp.187-194, 1999.
- [18] J.Nielsen, : “Noncommand User Interfaces”, Communications of the ACM, Vol.36, No.4, pp.83-99, 1993.
- [19] 大和正武, 神代知範, 門田暁人, 松本健一: “視線・マウス併用型インタフェースのドラッグ& ドロップ操作への適用”, 情処学論, Vol.44, No.1, pp.166-175, 2003.
- [20] 松本吉央, 怡土順一, 竹松憲太郎, 小笠原司: “リアルタイム顔・視線計測システムの開発と知的インタフェースへの応用”, 情処学論, Vol.47, No.7, pp10-21, 2006.
- [21] 米澤朋子, 山添大丈, 内海章安部伸治: “視線検出環境による擬人的媒体の階層的視線コミュニケーション”, 映情学技報, Vol.31, No.23, pp.1-4, 2007.
- [22] 石井亮, 大古亮太, 中野有紀子, 西田豊明: “視線と頭部動作に基づくユーザの会話参加態度の推定”, 情処学論, Vol.52, No.12, pp.3625-3636, 2011.
- [23] 阿部清彦, 佐藤寛修, 大山実, 大井尚一: “視線による重度肢体不自由者向けコンピュータ操作支援システム”, 映情学誌, Vol.60, No.12, pp.1971-1979, 2006.
- [24] 相田厚生, 三宅貴士, 森若誠: “視線入力システムによるポイント操作時のクリック方法とドラッグ& ドロップ方法の提案”, 人間工学, Vol.47, No.1, pp10-19, 2011.

- 
- [25] H.Istance, A.Hyrskykari, L.Immonen, S.Mansikkamaa and S.Vickers: “Designing Gaze Gestures for Gaming: an Investigation of Performance”, ETRA '10, ACM, pp.323–330, 2010.
  - [26] “World of warcraft”  
<http://us.battle.net/en/int?r=wow>
  - [27] H.Drewes, A.D.Luca and A.Schmidt : “Eye-Gaze Interaction for Mobile Phones”, Mobility Conference, ACM, pp.364–371, 2007.
  - [28] R.J.Krauzlis “The Control of Voluntary Eye Movements: New Perspectives”, The Neuroscientist, Vol.11, N0.2, pp.124–137, 2005.